Modelling and prediction of parameters affecting attendance to the NHS breast cancer screening programme

Arochena, Hisbel

Award date:
2003

Awarding institution:
Coventry University

MODELLING AND PREDICTION OF PARAMETERS AFFECTING ATTENDANCE TO THE NHS BREAST CANCER SCREENING PROGRAMME

H.E. AROCHENA

A thesis submitted in partial fulfilment
of the University's requirements
for the Degree of Doctor of Philosophy

2003

Coventry University
In collaboration with
Warwickshire, Solihull and Coventry Breast Screening
Unit

Table of Contents

Page.

Figures

Page

Tables

Abstract

This thesis focuses on the modelling and prediction of factors affecting attendance to screening invitations of the NHS Breast Screening Programme. The analysis is based on data collected by the Warwickshire, Solihull and Coventry Breast Screening Unit from 1989 up to 2001 with respect to invitation to screening for the prevention of breast cancer in non-symptomatic women.

Using a novel approach to the analysis of the data, from the perspective of the screening episode of each woman, rather than the usual analysis from the perspective of the screening round of the units, a statistical analysis is carried out on the whole registered population for the first time. Amendments to the current formulae for coverage calculations, the introduction of a new parameter (invitation rate) and the proposal for a reduction of the invitation period (period of time between two consecutive invitations) follows from the analysis.

A preliminary analysis of predictive methodologies, including traditional statistical methods and artificial intelligent methods, gives the foundation to the formulation of two new algorithms; the first, for the prediction of attendance of women to screening invitations, and the second for the prediction of occurrence of screening variation (change of appointment dates) of women to invitations. Both algorithms are based on neural network generated models able to learn from the previous screening behaviour

history of the woman, a technique not previously explored for the prediction of attendance.

The accuracy of the new proposed algorithm for the prediction of attendance to invitation is tested on a blind study using data not previously seen by the predictive system, and for which results were unknown at the time when the predictions were made.

From the obtained results, it is concluded to recommend the implementation by the NHS Breast Screening Unit of the two algorithms proposed for the prediction of the women's attendance and screening variation to their invitation for screening. With these predictions, women likely not to attend, or change appointment date, can be identified and appropriately targeted with the aim of increasing their attendance in the short term, and in the long term, reducing breast cancer mortality.

Acknowledgements

In bringing this thesis to life, and during the years of its development, there has always been support and understanding from colleagues, family and friends. This encouragement has been essential, therefore, I would like, in this section, to thank those who helped, in one way or another, to bring this project to an end.

Special acknowledgements go to the staff at the Warwickshire, Solihull and Coventry Breast Screening Unit, particularly to Dr. Mathew Wallis and Margot Wheaton for their dedication and availability, and to Chris Marano for always being willing to help arranging meetings and transferring information.

Acknowledgements also go to Dr. Paul Turner and Dr. Peter Barker from the Solihull and Coventry Health Authorities, respectively, for their support in making available the deprivation scores to postal area mappings data, and to Dr. Lillian Somerville and Emma Morrish, from the West Midlands Cancer Intelligence Unit for their support in making available data with respect to the yearly populations.

Acknowledgement also goes to Dr. Sue Moss from the Institute of Cancer Research for her early comments on the research.

I would also like to thank Mr. Ben Stonehouse from the SPSS group for providing me with the references to the Clementine algorithms implementations.

A very special thank you to Dr. John Gasser for his tremendous assistance with proof reading the initial drafts, specially his English corrections. Also to my supervisory team, Drs. Angela Whitford, Housh Mashhoudy, Alison Todman and specially my director of studies Prof. Raouf Naguib for his patience, encouragement and for always believing in me, even in my worst moments.

Acknowledgements go to my colleagues in the BIOCORE group, particularly for their good sense of humour.

A special thank you goes to my parents and friends for their financial and moral support, and especially to my daughter, Paulette, for her understanding and patience in sharing mum with her other "baby" (this thesis).

This project[1] has been generously jointly financed by the Warwickshire, Solihull and Coventry Breast Screening Unit in conjunction with the School of Mathematical and Information Sciences at Coventry University.

[1] As a project dealing with patient data, this project is registered with the University Hospitals Coventry and Warwickshire, NHS Trust, under Project ID 2000/112.

Glossary

Given the number of variables involved in the present work, a separate explanation of their definitions would be of assistance. There are two groups of variables, those relating to each woman and those variables defined for all women.

The variables repeated in each screening episode are defined in the same way for each one of them. A number added at the end of the name will specify to which episode the variable refers, e.g. FOA1, FOA2, etc.

The present work involves also two different ways of defining the screening year, one which ends on the 31st of March (current mode of measurement), and another, which ends on the 31st December. The annual variables are defined in the same way in both cases. A suffix in the name (D for December and M for March) will show which mode of measurement the variable refers to (Eg. NuminvD or NumscrM).

Ageband

Categorical variable assuming the following values;

1	< 50 years old at invitation
2	50 –54 years old at invitation
3	55-59 years old at invitation
4	60-64 years old at invitation
5	> 64 years old at invitation.
(0	if no invitations have been made)

AgeFOA

Age at First Offered Appointment (FOA). Number of years between DOB and FOA (measured as an integer).

Att (Attendance)

A categorical variable describing a woman's attendance. It assumes the following values:

1	invited but did not attend
2	invited and screened
(0	if no invitation)

Cancer

Binary variable

0	no breast cancer detected in the screening episode
1	breast cancer detected

Coverage

Proportion of women screened to eligible population in a given year, normalised to values between 0 and 1.

DOB

Date of birth. This is a date variable. (Note that in SPSS, all date formats are internally stored as the number of seconds from October 14, 1582.)

DOS

Date of first screening for the episode (SPSS-recorded date variable).

Early recall

An intermediate screening invitation occurring earlier than the routine screening interval.

Episode

Denotes the period of time elapsed between the first invitation of a particular woman up to the final outcome of her screening for that invitation.

Falsep

Binary variable

0	no false positive result in the previous episode
1	false positive result in the previous episode

FOA

Date of first offered appointment. (SPSS-recorded date variable)

Histcan

Binary variable

0	no previous history of breast cancer detection
1	previous history of breast cancer detection

HistFp

Binary variable

0 no previous history of false positive results

1 previous history of false positive results

Interval cancer

Cancers detected between screening episodes.

Intslip (Screening variation by intervals)

This is an integer variable that takes values depending on the *screening variation* value (women changes of appointments) as follows:

Table I. SCREENING VARIATION BY INTERVALS (INTSLIP) DEFINITIONS

Value	If appointment date is	Days	With respect to the invitation date
-8	more than three years	>1000	before
-7	more than one year but up to three years	[357 – 1000]	before
-6	more than six months but up to one year	[181 – 356]	before
-5	more than three months but up to six months	[91 –180]	before
-4	more than one month but up to three months	[31 – 90]	before
-3	more than two weeks but up to one month	[15 – 30]	before
-2	more than one week but up to two weeks	[8 – 14]	before
-1	one or more days but up to one week	[1 – 7]	before
0	no screening variation	0	on the day
1	one or more days but up to one week	[1 – 7]	after
2	more than one week but up to two weeks	[8 – 14]	after
3	more than two weeks but up to one month	[15 – 30]	after
4	more than one month but up to three months	[31 – 90]	after
5	more than three months but up to six months	[91 – 180]	after
6	more than six months but up to one year	[181 – 356]	after
7	more than one year but up to three years	[357 – 1000]	after
8	more than three years	>1000	after

Invitation period

Time elapsing between two consecutive screening invitations. (At present is three years)

Invitation Rate

Proportion of women invited to eligible population in a given year, normalised to values between 0 and 1.

KC62

National Statistics (return from the Screening Office).

KC63

National Statistics (return from the Health Authority).

Lengbs
(Round Length by British Standards)

[FOA (episode $_{x+1}$) – DOS (episode $_{x}$)]

Also known as slippage, this is a discrete variable which measures the difference in years between the date of last screening and the date of next first offered appointment. It is calculated as the difference in days divided by the number of days in a year, and rounded to the nearest integer.

Non-eligible women

Women excluded from the system because of either bilateral mastectomy or terminal illness.

Numinv
(Number of women invited)

Number of women invited in a given time period.

Numscr
(Number of women screened)

Number of women screened in a given time period.

Numtest

Number of tests women undergo in a single screening episode before final result for the episode. This variable is generated from the SEC.

Population

Number of eligible women in the population (aged between 50 – 64 years old) in a particular year.

Postanum

Integer equivalent to the three first characters of the Postal Code (for convenience). Thus, it is a categorical non-ordered variable. The correspondence is:

1	B10	60	CF7	119	NN1
2	B11	61	CH2	120	NN4
3	B13	62	CO4	121	NN5
4	B14	63	CV1	122	NN6
5	B15	64	CV2	123	NN7
6	B17	65	CV3	124	NR1
7	B23	66	CV4	125	NR2
8	B24	67	CV5	126	NR3
9	B25	68	CV6	127	OX1
10	B26	69	CV7	128	OX2
11	B27	70	CV8	129	OX7
12	B28	71	CV9	130	PL2
13	B32	72	DE1	131	PO1
14	B33	73	DE2	132	PO3
15	B34	74	DE6	133	PO4
16	B36	75	DH3	134	RG4
17	B37	76	DY1	135	S40
18	B38	77	DY6	136	S45
19	B40	78	DY8	137	SL6
20	B45	79	EH6	138	SN1
21	B46	80	EX1	139	SO3
22	B47	81	EX2	140	SO4
23	B48	82	EX3	141	SY4
24	B49	83	FK9	142	SY5
25	B50	84	FY1	143	SY8
26	B67	85	FY8	144	TA8
27	B72	86	GL1	145	TA9
28	B73	87	GL2	146	TF1
29	B74	88	GL5	147	TF2
30	B75	89	GL6	148	TN5
31	B76	90	HP3	149	TQ2
32	B77	91	HR2	150	TQ9
33	B78	92	HR3	151	TR1
34	B79	93	HR6	152	TR2
35	B8	94	HR7	153	TS7
36	B80	95	HR8	154	UB7
37	B9	96	HR9	155	W11
38	B90	97	IG1	156	W3
39	B91	98	IP2	157	WA1
40	B92	99	KA1	158	WA4
41	B93	100	KY7	159	WR1
42	B94	101	LD2	160	WR2
43	B95	102	LE1	161	WR3
44	B96	103	LE2	162	WR4
45	B97	104	LE6	163	WR5
46	B98	105	LE9	164	WR6
47	BA1	106	LL1	165	WR7
48	BA2	107	LL2	166	WR9
49	BH1	108	LL4	167	WS1
50	BH2	109	LL6	168	WS4
51	BH4	110	LN1	169	WS7
52	BN1	111	LN5	170	WV1
53	BS1	112	LS1	171	YO1
54	BS2	113	MK1		
55	BT8	114	MK4		
56	CA1	115	NE6		
57	CB1	116	NG1		
58	CB2	117	NG2		
59	CB3	118	NG3		

Roundle (Round length)

[FOA (episode $_{x+1}$) – FOA (episode $_{x}$)]

This variable measures the difference in years between the date of first offered appointment of one screening episode and the following one. It is calculated as

the difference in days divided by the number of days in a year and rounded to the nearest integer.

Screening End Code (SEC)

Categorical variable. It is a compilation of the outcomes of all the tests undergone by a woman in the screening episode. Table I gives the nomenclature used for the coding of the screening end code.

Table II. SCREENING END CODE (SEC) DEFINITIONS

Classification	Screening (S)	Assessment (A)	Cytology (C)	Core biopsy (W)	Histology (H)
Normal outcome (-)	S-	A-	C-	W-	H-
Abnormal outcome (*abn*)	S*abn*	A*abn*	C*abn*	W*abn*	H*abn*
Cancer Suspected Positive outcome (+)	S+	A+	C+	W+	H+

Screening variation

Difference in days between date of first offered appointment (FOA) and date of screening (DOS) for a given episode. It may assume negative and positive values: negative for days preceding FOA, and positive for days following FOA. Zero values imply no screening variation. However, no values will be assumed if neither invitation nor screening for the episode has occurred.

Sx

A unique identification number for each individual. This is a numerical qualitative ordered variable, and gives information about how early or late a woman enters the system. The later the entry the higher the number.

Townsend Deprivation Score

Measurement of socio-economic deprivation. It measures the level of deprivation based on housing, income, car ownership and educational level amongst other socio-economic factors. The higher it is, the more deprived the area.

Townsref

Integer variable grouping by interval the different values of the Townsend deprivation scores. It takes values:

-3 $-10 \leq Townsend \leq -8$ (richest areas)

-2 $-7 \leq Townsend \leq -5$

-1 $-4 \leq Townsend \leq -2$

0 $-1 \leq Townsend \leq 1$ (average areas)

1 $2 \leq Townsend \leq 4$

2 $5 \leq Townsend \leq 7$

3 $8 \leq Townsend \leq 10$ (poorest areas)

Type

Type of invitation for a given screening episode. This is a qualitative variable which can take two values:

First call

Routine recall

Typebin

Qualitative variable, which may assume the values:

1	First call
2	Routine recall
(0	No invitation for the episode)

Uptake

Percentage of women who, having been sent an invitation for screening in a given year, attend a screening unit and undergo mammography in response to that invitation. No allowances are made for returned letters or refusals. Normalised to take values between 0 and 1.

Chapter 1

Introduction

1.1 Introduction

Breast cancer is the most common cancer in women worldwide, accounting for approximately 30% of all female malignancies. Quinn *et al.* [1] reported in the late 1980's that mortality in England and Wales was not only higher than in most western European countries, but was amongst the highest in the world. Incidence, on the other hand, was similar to that in eastern European countries. In other words, survival was worse than in the rest of Europe. Although the reasons for such poorer survival rates remain unclear, there are suggestions that late diagnosis and poor adherence to treatment protocols may be contributing factors.

Over the period 1993-95, the United Kingdom was identified, through the Global Breast Cancer Death Rates Statistics [2], to have the third highest breast cancer death rate in the world despite having a population-based breast cancer screening programme since 1987. This UK standing is alarming, though it hides the fact that, since 1989, the crude death rate due to breast cancer has decreased from 53.4 to 44.6 per 100,000 in 1997, as documented in the WHO Databank [3].

In 1997 there were 33,100 new registrations of breast cancer in women in England and Wales: almost 30% of all cancers in women [1]. The lifetime risk of being diagnosed with breast cancer is almost 11% (1 in 9). Mortality has, however, declined since the late 1980's. It began to fall soon after screening started (in all age groups), and by 1999, overall mortality was approximately 20% lower than the level attained in the 1980's. About one third of this decline is attributed directly to screening, and two thirds to improved treatment by tamoxifen and chemotherapy and to indirect effects of screening such as raised awareness leading to earlier presentation and diagnosis outside the screening programme. Also, since the early years of screening, there have been substantial improvements in sensitivity measurements as a result of the increased use of two-view mammography, the use of higher film densities, and increasing experience of radiologists.

The NHS Breast Screening Programme was introduced in response to the recommendations of a working group chaired by Professor Sir Patrick Forrest [4] based on randomised controlled trials in America and Sweden. Its principal objective is to reduce mortality from breast cancer in the screened population. In order to achieve this objective, it targets to screen women aged between 50 and 64 every three years with the aim of detecting cancers at an earlier stage, when treatment is more effective. This should lead to a target reduction in mortality of 25% [5].

Continuous monitoring of the performance of the NHS Breast Screening Programme is essential. In the absence of a control population, a variety of proxy measures have been developed to evaluate the programme. The two principal factors identified as

affecting mortality reduction are improvements in cancer detection and population compliance.

Amongst the proxy measures developed to evaluate the programme, are rates of screening detection of cancers, interval cancers, cancers in non-attenders [6], cancers in non-invited women, slippage [7] and attendance. These measures show the efficiency of the programme; therefore it is necessary to study them rigorously in order to (a) assist in reducing mortality and/or morbidity, and (b) perform quality assurance evaluations of the programme.

Uptake and coverage have routinely been used to monitor the performance of screening units [8-10]. However, their underlying methods of calculation have varied over the years. It is therefore desirable to identify a unified model of measurement in order to carry out comparative studies and assist in forecasting their trends.

Other factors playing an important role in the quality assurance of the programme are methods of data collection and storage.

Numerous studies have been carried out in the UK and worldwide addressing some of the above issues. Notable examples are the follow-up of the Edinburgh trial in the UK [11,12], the systematic review of the UK Breast Screening Programme [13-17], critical assessments of efficacy [18-20], methods of screening [21,22], screening age [23] and various definitions (such as that for interval cancers [24]). Nevertheless, there are still questions to be answered and topics open to research. Topics such as the influence of attendance at previous episodes on the prediction of future attendance, the effect of

screening variation on uptake, and the improvement of the coverage formulation, are of current interest. These issues are addressed in this study.

Since statistical methods of analysis have been successful in measuring the validity and performance of screening programmes (trials and population-based), as well as in modelling parameters involved in the achievements of particular units or regions [25-31], these form the basis of an initial study. The subsequent predictive analysis is approached using artificial neural networks [32-39], rule induction algorithms [39], and decision trees [34,40,41]. These results are then compared with those of the traditional statistical logistic regression method [42-47], and represent an extension of existing work, within BIOCORE, related to clinical applications of intelligent computing in oncology [48-52].

1.2 Statement of the problem

As stated in the introduction, most research carried out worldwide in this area is based on control-based studies. The UK Breast Cancer Screening Programme is a population-based programme, and as a consequence, there is no control group to help in its assessment. This highlights the importance of proxy measurements in monitoring the effectiveness of the programme. Two of the most important measurements are *uptake* (measuring the performance of the invitation process) and *coverage* (measuring the efficiency of the programme in covering the whole population). However, both of these are aggregates of the *number of women invited* and the *number of women screened*, which are themselves aggregates of patient level variables. They are, therefore, overall measurements of the *attendance* of women at their invitation to screening.

The ways in which *uptake* and *coverage* have been measured have changed over time during the running the programme, but no assessment of the current formulae used has been performed, neither has there been any attempt to study their behaviour using a unified measurement.

The literature shows many studies based on the influence of socio-economic factors on *attendance* and *cancer death* outcomes worldwide, but very little has been achieved on the influence of the intrinsic factors of the screening process affecting attendance from one episode to the next.

Furthermore, very little research has been performed to analyse the influence that changing screening appointments has on the overall measurements of *uptake* and *coverage* over a given time period.

It is clear that early detection of malignancies is the most important factor in the battle to reduce breast cancer death, and the screening programme objective is to achieve this through mammography. However, very little can be achieved if women do not attend their appointments. Prediction of *attendance* in order to take decisive steps for improving the performance of the programme is a crucial matter. A statistical approach for forecasting *attendance* [53] was not successful, since it was based on aggregate data reported on the KC63 forms, and not from individual data. It is important to know not only the percentage of women who are likely to attend or not, but also to provide a tool that, given a particular woman and her screening history, will predict her future *attendance* and *screening variation.* With this information available, effective measurements could then be taken in order to increase the attendance of women at

screening. Furthermore, once the prediction for each woman is known, the *uptake* and *coverage* for a particular episode can also be predicted using their formulation in the predictive results rather than in the real data.

Classical predictive statistical methods, such as Logistic Regression have proved to be very efficient tools for medical predictions in other fields. Nevertheless, the complex relationship structure of categorical variables involved in forecasting *attendance* and *screening variation* points to the need for using more powerful tools, like Artificial Neural Networks (ANNs), rule induction and add-hoc heuristic techniques which have been shown in the literature to be useful in the prediction of other cancer-related features.

1.3 Aims

The results presented in this thesis use data from the first ten years (3⅓ rounds) of screening at The Warwickshire, Solihull and Coventry Breast Screening Unit (a total of 281,415 potential screenings for 147,432 women).

The project aims are to:

(a) statistically model the variables affecting *uptake* and *coverage*, (measures of the relationship between women invited, screened and the total eligible population);

(b) identify the variables affecting *screening variation*, *attendance* and *round length*, (measures of the elapsed time between when women are invited and when they are screened, involving current and previous episodes of screening);

(c) determine the validity of the formulae used to calculate *uptake* and *coverage*, and make recommendations for future improvements of these formulae to the Breast Cancer Screening Programme;

(d) develop intelligent computing techniques for predicting *attendance* and *screening variation* patterns at patient level, which will naturally lead to the prediction of *uptake* and *coverage* at aggregate level.

1.4 Structure of the thesis

The thesis is structured as follows.

Chapter 1 presents a statement of the problem and outlines the aims of the study. A review of related current literature is carried out in Chapter 2, meanwhile the data acquisition and preparation is explained in Chapter 3.

Chapter 4 concentrates on the statistical descriptive analysis. In particular, a sub-section is dedicated to the analysis of *uptake* and *coverage* focusing on the formulation of the revised formula proposed for measuring *coverage*, and presents other recommendations for the Breast Screening Programme. The significance of the results obtained by the statistical analysis is also discussed.

Chapter 5 introduces the artificial intelligence approach to predictive analysis. It focuses firstly on the analysis of the relationships and predictors of *attendance* and *screening variation*, and secondly on a description of the data and methods used to carry out the predictive analysis.

Chapter 6 concentrates on the prediction of *attendance* using AI methodology compared with a well established statistical method as a benchmark. Formulation of a new algorithm for the prediction of *attendance* is presented as the closing section of this chapter.

In Chapter 7, a similar approach to the one described in the previous chapter is used, but this time in relation to the prediction of *screening variation.* A consequent extension to the algorithm introduced is developed.

Chapter 8 concentrates on a blind study focused on the prediction of *attendance* to the screening invitation using the proposed predictive algorithm. The discussion of the results obtained in the project and the conclusions drawn from these are presented in Chapter 9.

Appendices A and B give the data tables and a graphical description of the impact of the *screening variation* per episode. Appendices C and E present the data tables of the predictive methods for *attendance* and *screening variation* respectively. Appendix D, on the other hand, shows the Decision trees obtained for the prediction of *attendance* and the decision rules for *non-attendance*. In Appendix F the predictive algorithm results for

the attendance of women to the screening invitation for the blind study are given. An exhaustive explanation of the classical statistical and predictive methods used can be found correspondly in Appendices G and H. Appendix I contains photocopies of the published papers.

1.5 Dissemination of results

Throughout the lifetime of this research, a number of publications have been disseminated to the specialist research community.

Seven reviewed conference papers have been published, two of them in 2000 [54, 55], two in 2001 [56, 57], two in 2002 [58, 59] and one in 2003 [60]. In addition, two papers have been published in internal proceedings in Coventry University [61, 62]. A full report [63] has been submitted and approved by the sponsoring collaborating body, the NHS Breast Screening Programme.

Chapter 2

Literature Review

Following the statement of the problem outlined in Chapter 1, and the aims of our project, this chapter presents a review of the literature relevant to the development of the thesis.

In this review, two main approaches need to be taken. The first is a review related to screening programmes, their development, analysis of performance and factors affecting them. The second approach deals with methodologies used to analyse medical data, specifically medical data mining and predictive methods applied in medicine, including the traditional statistical methods and the emerging use of artificial intelligence methodologies.

2.1 Breast cancer screening programmes

2.1.1 General overview

As analysed by Forrest and colleagues in the report which laid the foundations for the implementation of the UK NHS Breast Screening Programme [4], up to 1986, numerous studies evaluating the effect of screening on breast cancer mortality were carried out worldwide. The successes reported by such studies encouraged the introduction of

organised population based breast-screening programmes in several countries, including the UK.

As part of the steps towards the implementation of such programmes, three main subjects were discussed [8]:

- the *methods of screening:*
    - Clinical Breast Examination (CBE) [19, 22]
    - Breast self-examination (BSE) [19, 22, 64]
    - Mammography [5, 6, 11, 12, 20, 22, 25, 26, 27, 64-74]

Screening with mammography being the most preferred methodology implemented.

- the guidelines for the *lower age limit* of mammography screening [19]:
    - Younger than 40 years old [75]
    - 40 –49 years old [23, 64, 68, 75]
    - 50 years old [64, 68]

50 years old being the most widely implemented.

- the *systems to recruit women to screening:*
    - Invitation letter [6, 76]

- Invitation by phone
- Others

Invitation letters are the most widely reported implementation for first invitation to screening.

All of the above subjects may be grouped under a main title, *quality assurance activities,* which are an evaluation priority for all the screening programmes [73], and for the IBSN (International Breast Cancer Screening Network) at world level.

Particularly for the UK, several publications have been disseminated relating data tables, definitions and guidelines to methods of evaluation of the quality of the breast screening programme [77-79]. In addition, annual reviews of the UK NHS Breast Screening programme have been carried out since 1993 [13-17], [80].

Evaluation of the screening programme performance has been mainly focused on measures such as:

- Death rates [5, 11, 12, 20, 21, 23, 26, 64-68, 81]
- Uptake [6, 71, 73, 82-87]
- Coverage [13-17]

Although the *coverage* formulation has been well defined and accepted worldwide, the methods by which the data is collected for its calculation vary from one country to another. In the UK, even though the data collection methodology has varied over the

years, no unified analysis taking the same approach of data collection has been reported. Moreover, no assessment of the coverage formulation has been reported to date, either.

Uptake, as another proxy measurement of the programme performance, has been measured since the introduction of the programme in the UK (1989), monitored and reported annually by each screening unit and by the Breast Screening Programme nationally. It is a measure of how well the programme is accepted by those women invited. Nevertheless, no measure assessing how effective the programme is at inviting women has been defined. This may be due to a general belief that this issue is covered by the *coverage* formulation, but the latter parameter only measures the proportion of the population covered by screening, and assumes that those not screened are not attendees. A question arises then: could there be a percentage of this population that is not reached by the invitation process in the first instance?

2.1.2 Factors influencing acceptance and participation

Having accepted that screening programmes are a valuable means in the fight against breast cancer, the need for a closer look at the mechanisms of their operation becomes an important issue. The most important factor in determining the potential success of breast cancer screening programmes is the positive response of women to it.

Increasing the response of women to breast screening and the possible factors influencing their *attendance* have been a subject of great interest to researchers

worldwide. Studies related to this issue have been carried out in most countries where some form of screening programme has been implemented.

In a broad sense, *uptake* is nothing more than an aggregate of the women's *attendance* to screening. Thus, in order to influence positive changes in the *uptake*, the acceptance and participation of women to screening invitations ought to be analysed.

Attendance of women to screening programmes has been extensively studied worldwide through full population programmes and / or control based studies. Amongst the main factors reported as influencing the *attendance* of women to screening, the following can be mentioned:

- age group [76, 82, 88]
- previous health behaviour [22, 72, 85, 89, 72]
- previous screening participation [22, 25, 72, 86, 89]
- past uncomfortable experience of breast screening [90]
- minority groups [69, 83, 89, 91]
- socio-economic factors [70, 72, 82, 84, 85, 87-72, 92, 93]
- previous cancer history [72]
- distance to the screening facilities [76, 85, 88, 92]
- density of medical amenities [88]
- intrinsic factors associated with the programme [20, 25, 26, 85]
- type of invitation [6, 87]
- suffering from other medical problems [72, 82]

2.1.3 Attendance and social deprivation

Socio-economic factors, and in particular poverty, are reported as having a significant influence on the *attendance* of women to screening [88].

Studies carried out in the UK which investigate parameters affecting *uptake* and their relationship with socio-economic factors concluded that variation in the *uptake* of breast cancer screening is closely related to *social deprivation* [84].

Several indices have been developed which aim at explaining the levels of poverty in different contexts. In the UK, the most common indices reported are,

- Jarman Index
- Carstairs Index
- Townsend Deprivation Score
- LWT Breadline Britain Index (LWT)
- Index of Local Conditions (ILC)

Based on the apparent influence that deprivation has on health-related outcomes, most of the social-deprivation analyses are based on the *Townsend deprivation score* [94]. The selection of this particular score has its foundations in the results obtained through a UK-based research by Payne, Payne and Hyde [95]. Although primarily directed at sociologists, this study analyses the different indices of poverty and *social deprivation*, highlighting the importance of *social deprivation* indices in society, in general, and how it operates for different groups. One important conclusion reported in this work with

relevance to ours, is that the *Townsend deprivation score*, out of all the other mentioned social-deprivation indices, is the one that relates more closely to all the other indices. In other words, the *Townsend deprivation score* groups in itself the factors taken into account by the other indices. This conclusion gives, as such, strong justification for the use of this particular score when analysing the effect of *social deprivation* on *attendance* to breast cancer screening.

Nevertheless, the use of this score needs to be carefully monitored, because, as reported by Collins *et al.* [96], Townsend score can report higher values than reality when individual addresses are used to match enumeration districts via the address postcode. It is suggested that for problems involving resource allocation and for research into relationships between health outcomes or service *uptake* and *deprivation*, it may be necessary to quantify the level of error introduced through using postcode to enumeration district matching. Focusing particularly on breast cancer screening in the UK, a more appropriate matching of individuals' addresses to *deprivation scores* should be developed by the units when taking this factor as a predictor for *attendance* [87].

A possible solution to this problem, not discussed in the literature, could be to use intervals of values of the *Townsend deprivation score* (discretisation of the variable), instead of using the original continuous values.

2.1.4 Delayed attendance and slippage

Although the subject of screening appointment changes by women is raised in the literature, no analysis of the influence of these factors has been reported for non-symptomatic women. Neither has a study of the effect that these appointment changes can have on the programme performance been reported. As a direct consequence, no attempt to predict women's change of appointments has been reported.

Nevertheless, studies carried out focusing on symptomatic women have analysed factors predicting delayed presentation to screening [97]. Amongst such factors, those that could be equally relevant to non-symptomatic women are:

- Age group
- Marital status
- Education level
- Ethnicity, minority groups

Delayed *attendance*, on the other hand, is closely related to screening *round slippage* (rounds longer than 3 years) and *interval cancers* (cancers detected between two 3-year interval screenings). These two factors are publicly recognised to affect breast cancer screening programmes [7, 90, 98].

Round slippage by itself, on the other hand, can have an impact on the increment of *interval cancers*. Although Threlfall *et al.* [99] proposed, as a way of decreasing *interval cancers*, to offer younger women in the programme screenings at less than the actual three years interval.

A more radical suggestion was proposed by Faux *et al.* [24]. They proposed that the definition of *interval cancers* should be changed to include cancers arising at 36 months (3 years) or more from the last screening. Their conclusion was that the exclusion criteria used in the present definition of *interval cancers* had a significant impact on observed *interval cancer* rates.

Threlfall et al. stated:

"*The occurrence of some interval cancer is inevitable, but a high rate of interval cancers may indicate poor sensitivity or an unsuitable long screening round, or both*". [99]

They also highlight problems arising due to the programme structure of invitations. In reality, all women are not screened at exactly three yearly intervals as it is assumed, but rather in longer or shorter time periods. They concluded that the exclusion of those cancers from *interval cancer* data masked the problem of late re-invitation. They fail to point to another problem masked under the same umbrella, the problem raised by the women's changes in dates of screening appointments (delayed *attendance*).

2.1.5 Interventions to prevent delayed attendance and non-attendance

There is a recognised believe that there exist significant differences between previous participants and non-participants [6, 25], and that determination of intentions and past screening behaviour could be used to improve participation and adherence to breast

cancer screening programmes. Once those women likely not to attend or incur a delayed presentation are identified, appropriate interventions could be taken to avoid or minimise this behaviour.

Several typical actions have been identified and reported. Amongst those are:

- re-invitations [6, 71, 72, 87]
- general practitioner (GP) letter addressing [71, 83]
- reminder letters [70, 71]
- phone calls [70]
- physician recommendation [22, 70, 71, 72, 83, 86, 100]
- cultural relativistic approach [69, 83, 86, 91]
- clinic based interventions [70, 90]
- patient education [70, 71, 86]
- increase staff [70]
- offer of transport to the screening centre [83]

2.2 Prediction of attendance

Identification of predictive factors of *attendance* to screening has been discussed and reported extensively in the literature. Amongst the predictors of *attendance* to screening has been cited,

- physician recommendation [22, 72, 87, 100]

- age group [22, 82, 88]
- household income [22, 70, 82]
- deprivation [82, 87, 88]
- previous screening behaviour [6, 25, 72, 86]
- intention of participation [25, 86]
- consequences of screening [25, 100]
- marital status [72, 82, 86]
- employment status [71]
- previous experience of cancer [72]

However, although such identification has been achieved, no attempt to predict attendance from patient level information has been carried out.

2.2.1 Methodology overview and predictive analysis approach

The methodology used for prediction in the medical literature can be subdivided in,

- social cognition models [86, 100, 101]
- statistical methods [25, 72, 73-75, 86, 87, 100, 102-108]
- fuzzy logic methods [51, 52, 109]
- rule induction and genetic algorithms [109-111]
- artificial neural networks [49, 50, 51, 52, 102-105, 107-110, 112, 113, 124-127]

The most reported social cognition models are:

- The Health Believe Model [86, 100]
- The Theory of Planned behaviour [86, 101]

However, the data available for this study do not allow a successful implementation of these approaches as it does not include sufficient socio-economic information relating to individual women.

Consequently, this review will concentrate on statistical and artificial intelligence methods.

2.2.1.1 Statistical methods

Statistical methods are well recognised by the scientific community, and especially well established amongst the medical community. In particular, they are recognised as successful tools in medical applications for the prediction (and / or prognosis) of different medical features. The most reported statistical methods used for these purposes are:

- Logistic regression [25, 51, 52, 87, 72, 102, 105, 124]
- Markov chains [74, 75]
- Cox regression [102-104, 106-108]
- Principal components analysis [105]
- Factorial analysis

- Discrimination analysis [109]
- Bayesian likelihood analysis [73]

Most of the studies in the previous sections, used statistical methods in their design and analysis.

Statistical methods, successfully applied in medical research, are documented in various reference books, which give exhaustive explanations of them. Particularly relevant discussions of the most commonly used techniques can be found in Armitage [28], Puri [29], Altman [30], Bland [31], Collet [41], Agresti [42, 43], Cox [44], Hosmer & Lemeshow [46], McCullagh & Nelder [47], Cooper [114], and Everitt [115]. These books are also mentioned as references for computational implementations of the methods in well-established statistical software packages such as SPSS, Clementine and Unistat.

2.2.1.2 Artificial intelligence methods

More recently, artificial intelligence methods (AI) have emerged as potentially powerful tools in medical applications.

However, the introduction of new techniques in a well-established area is always controversial. In defence of the introduction of the use of Artificial Neural Networks (ANN) in Biomedicine, Burke [112] expressed a main paradigm:

"It is difficult for traditional statistical methods to capture complex systems because traditional methods attempt to find the model that best fits the statistician's understanding of the phenomenon; complex systems are difficult to understand and therefore difficult to fit with a simple model. Artificial neural networks are nonparametric regression models. They capture any phenomena, to any degree of accuracy (depending on the adequacy of the data and the power of the predictors), without prior knowledge of the phenomena. ... Artificial neural networks are a powerful method for capturing complex phenomena, but their use requires a paradigm shift, from exploratory analysis of the data to exploratory analysis of the model"

Several reference books are reported as classics for AI methods and algorithms, e.g. Goldberg [39]. More specifically, Patterson [32], Ripley [33], Berry & Linoff [34], Bigus [35], Bishop [36], and Masters [37, 38] deal with ANN, while Berry [34], Mitchell [40] and Quinlan [41] deal with Rule induction and Decision trees methods. Many of these references are also cited in the documentation of algorithms implemented as part of Clementine and several libraries of Matlab.

Rule induction techniques and genetic algorithms applications have been used and recommended when generating classification rules and are particularly useful in carrying out exploratory analysis of generated models. However, the dimensionality occurring when dealing with large databases involving a large number of features is a common problem encountered [109].

In particular, applications of ANN have been reported, as often achieving better results than the statistical methods, in

- Survival analysis [102-105, 107, 113, 124, 125]
- Medical prognosis [102, 104, 105, 107, 112, 125, 126]
- Treatment outcome [103, 108]
- Disease classification [49, 109, 110, 127]

amongst others.

Nevertheless, when applying ANNs in medical data some common pitfalls should be avoided. Schwarzer *et al.* [116], provide a good discussion of current approaches and their comparison with statistical methods (particularly Logistic regression). Frequent criticisms of applying ANN methods are:

- Doubtful convergence to the real function that they are approximating in cases of small data
- Careful selection of the sample size needed in order to have an accurate estimation of misclassification probabilities
- Risk of over-training
- Non separation of the data into training, validation and test data sets
- Reporting results comparisons with inappropriate statistical methods

Although such points are normally raised by retractors of the use of AI methodology in oncology, the points raised are valid and need consideration when applying and reporting results obtained by means of the use of these approaches.

Concentrating on applications of ANNs in the medical environment; it can be said that, although an extensive variety of methods have been exploited, some of the most common ANN methods applied in the medical context are:

- Feed forward back propagation neural networks [51, 52, 102, 107, 108, 110, 112, 124, 125, 127]
- Radial basis functions [49]
- Pruning neural networks [113]

No reports could be found that discuss any of these methods in relation to the prediction of attendance to screening.

2.2.2 Prediction of attendance to full population based breast cancer screening programme

As demonstrated by the literature discussed in previous sections, prediction of *attendance* could clearly facilitate an improvement in the outcomes of the UK NHS Breast screening programme.

An attempt to predict *attendance* was carried out by Lancucki & Babb [53]. Using reported data from the UK NHS breast screening programme, they proposed the use of a new measure of *uptake* based on the *estimated cumulative attendance* (ECA), i.e. the prediction of the *attendance* as a cumulative frequency. They claim that this measure has the advantage of being a summary statistic that can be evaluated from year to year. This is the estimated percentage of total possible women-attendances which would occur if

the annual screening programme *uptake* rates continued at the same level (i.e. assuming no variation in the *uptake* rates). They also reaffirmed the statement supported by the screening programme that to reduce high mortality from breast cancer, both a high *cancer detection* rate and high *attendance* and *re-attendance* rates are required. However, ECA is an aggregate measurement and, as a result, it fails to predict the likeliness of *attendance* to the invitation by a particular woman. This is important in targeting intervention activities, instead of just "hope" for the fulfilment of the hypothesis of maintaining the same *uptake* rates [64].

No reports could be found of algorithms or methodologies designed to predict the *attendance* of a particular woman for her breast cancer screening invitation given her previous screening history.

There is currently a lack of predictive tool for *attendance* to screening that needs to be addressed. Any attempt to develop such a tool should take into account the different methodologies available and present a comprehensive comparison of their performance before choosing a particular method for implementation. At this stage, another issue needs to be taken into consideration. Methods computationally implemented under different platforms are bound to suffer from biases due to implementation differences, so, as a first approach, a comparison under a well recognised environment for all methodologies is needed. Such a need has not yet been fulfilled as far as the literature report.

2.3 Main findings of previous work

Although the present work builds on the body of information highlighted in the previous sections of this chapter, some particular findings played a crucial role in the development of the project, as it stands.

The analysis of the aims and objectives of the Breast Cancer Screening Programme, as well as the need for such a programme, reported by Forest *et al.* [4] played an important role in the understanding of the problem and development of this work, as it provided the ultimate motivation for the project, namely, to improve early cancer detection in order to decrease death rates due to breast cancer. No applied research can be done without a sound understanding of the objective of such research; Forest's report, in justifying the need for a full population breast screening programme, contributed to achieve such understanding.

The decision to use the Townsend Deprivation Score in our work as the socio-economic factor influencing attendance was based on the results reported by Payne, Payne and Hyde [95]. Their work established that the Townsend score was, of all existing socio-economic indices, the one that best brings together all of the factors taken into account by the other indices. The way in which we use the index, however, (in ranges rather than in the value itself) is based on the results of Collins *et al.* [96].

As discussed in Section 2.1.3, mapping socio-economic factors introduces the possibility of errors due to the variability of such factors even in small geographical areas. Collins *et al.* suggest that this index could report higher values than in reality

when mapped using individual addresses to enumeration districts via the address postcode (mapping used in our project due to the lack of individual information).

Bearing in mind the recommendations of Gatrell *et al.* [87] while working in breast cancer screening in South Lancashire, of the need for development of a more appropriate matching of individual's addresses to deprivation scores, the decision was made to use intervals between values of the Townsend deprivation score in our work rather than the particular index value, thus reducing potential mapping errors.

Detection of screening variation as a factor affecting the screening programme and the decision to develop a means of predicting it in our project, was based in the results obtained by Threlfall *et al.* [99] and Faux *et al.* [24] while looking at the impact of delayed attendance and round slippage on the appearance of interval cancers.

Furthermore, the approach of looking at the prediction of attendance and screening variation from the point of view of each woman over her screening history, (rather than as an aggregate), is based on the results reported by Stead *et al.* [6] and Lechner *et al.* [25]. These studies recognise the existence of significant differences between previous participants and non-participants and point towards the women's past screening behaviour as a possible way of increasing their participation and adherence to screening.

In order to carry out the development of predictive algorithms, the identification of possible predictor factors was needed. The results obtained in our work, relating to the identification of predictors of attendance, corresponds with the predictors identified

and reported in previous works, including those of Stead *et al.* [6], Solomon *et al.* [22], Lechner *et al.* [25], Paskett *et al.* [70], Lagerlund *et al.* [72], Edwards and Jones [82], Bish [86], Gatrell *et al.* [87], Challier *et al.* [88], and McDonald [100].

The decision to use Artificial Neural Networks (ANN) as the predictive methods, was influenced by the work carried out by Burke [112]. His proposed paradigm shift, from exploratory analysis of the data into exploratory analysis of the model provided the intellectual motivation for our prediction approach. Their particular use on breast cancer related prediction was also encouraged by Ripley's work [102]. IN applying the techniques, the work of Schwarzer *et al.* was carefully considered in order to avoid common pitfalls, such as: the use of large data sets to increase convergence credibility; careful selection of sample sizes to guarantee accurate estimation of the misclassifications probabilities; the use of techniques to avoid over-training of the models; careful separation of the data into training, validation and tests data sets; and the appropriate comparison with an equivalent statistical method.

The use of the Logistic Regression method as a comparison baseline of the ANN results followed from the results reported by Lechner *et al.* [25], Seker *et al.* [51], Lagerlund *et al.* [72], Gatrell *et al.* [87], Ripley [102], Burke *et al.* [105] and Intrator and Intrator [124] in applying this particular method.

Finally, the decision to use Rule induction techniques rather than Genetic algorithms in the exploratory analysis of our generated predictive models was founded on the results of Ponsan [109] and its recommendations when working with large databases, as it was our case.

From our research of the literature, only one previous work was found that considered the prediction of attendance to breast screening. The work of Lancucki and Babb [53], although pointing out the need for prediction of attendance, follows a very different approach to the problem than ours. They predicted attendance from aggregate data and consequently, can only predict the number of women who will, or will not attend. They are unable to predict an individual woman's behaviour, as addressed by our work. The possibility of achieving this latest result is developed in our work.

Chapter 3

Data Acquisition and Preparation

3.1 Preliminaries

3.1.1 The invitation process

The National Health Service Breast Screening Programme works to a 3-year cycle and aims to invite all eligible women every 3 years, with the first invitation taking place before a woman's 53rd birthday.

Screening may be performed at either static or mobile sites. Women are invited on the basis of GP practice and locality, and a screening plan is devised to ensure that all women are invited/reinvited within the 3-year cycle. This system was set up as a result of evidence showing that inviting by GP practice and with GP support will boost *attendance* rates [71]. A mobile unit placed within the geographical area being screened also promotes *attendance* [92].

Approximately 3 months prior to the screening process starting, the screening office specifies the 'batch' of women due to be screened by means of age and GP. This information is transmitted to the Health Authority system which in turn selects all eligible women and downloads their corresponding details to the screening office

computer system. The screening office computer system subsequently matches the women identified against the existing database to further identify women who already have an *Sx* number (unique identifier of the women in the screening database) and then allocates a number to all new women. At this point, an individual woman can be accidentally allocated a second number (e.g. due to change of name or a transcription error in NHS number or date of birth). Once the 'batch' has been accepted and created on the screening office system women are allocated a timed appointment.

Initially the *Sx* number was allocated in a sequential manner. Since 1992, when two-view mammography of each breast was introduced for all new women, and women attending a second or subsequent round were offered one view [27], numbers have been allocated randomly. This is achieved through a bespoke programme that ensures that radiographers take the correct number of views.

From the above, it follows that although GP practices retain the same order in each 3 year cycle an individual woman will not maintain the same relative position within the batch.

Fig. 3.1 illustrates the actual invitation process in the Warwickshire, Solihull and Coventry Breast Screening unit.

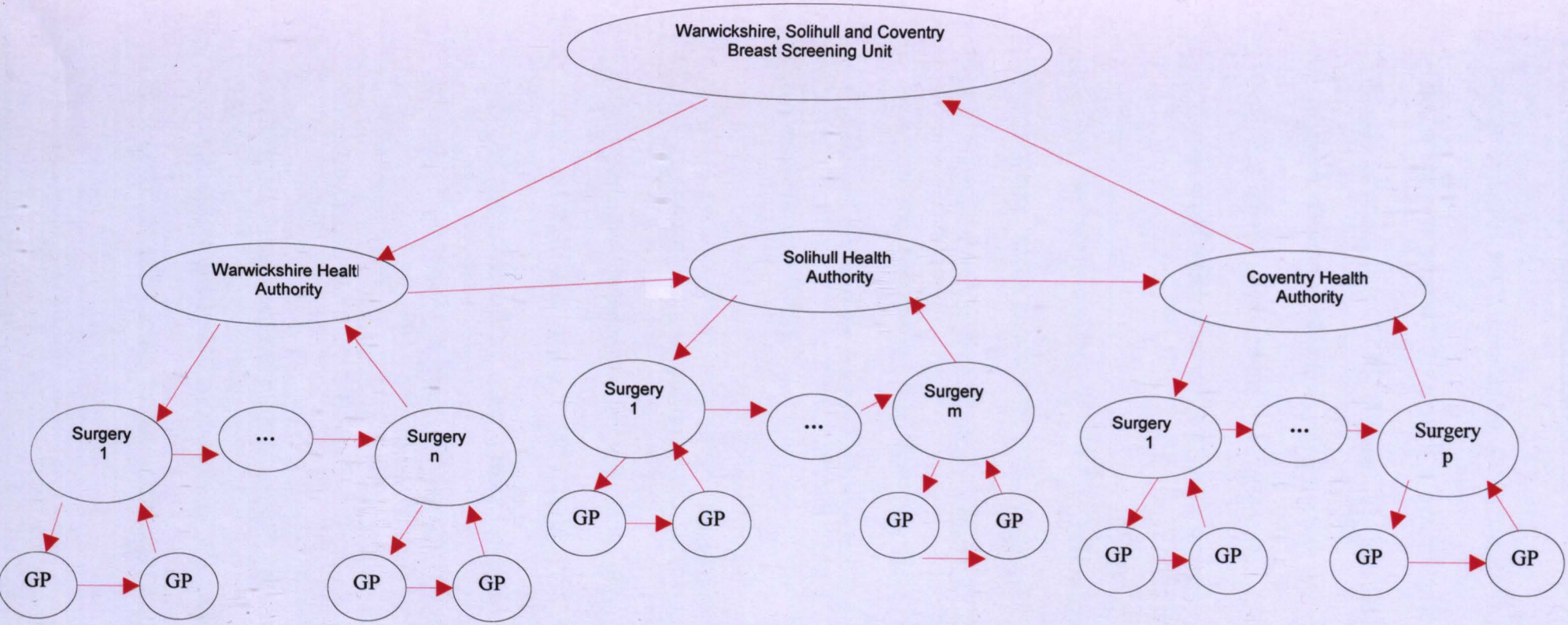


Fig. 3.1 Flow of the invitation process

3.1.2 Data preparation

The raw data provided by the Warwickshire, Solihull and Coventry Breast Screening Unit consisted of the first ten years (3⅓ rounds) of screening within the unit (a total of 281,415 potential screenings for 147,432 women). The variables included in the initial data (only variables relevant to this study are recorded) are the *Sx* number, *date of birth*, *date of first offered appointment*, *type of invitation*, *date of screening* and *screening end code*, all of which were stored for each single invitation.

The original data was transferred from a MUMPS database held by the unit into a Windows based system, which involved the initial conversion of the data into a text format in order to export it into its final Windows compatible format. Microsoft Excel applications routines, implemented in Visual Basic and SQL, were used to reorganise the data and validate it. Some of the macros used were based on previous work within the BIOCORE research group [117].

Throughout this data transfer process, a problem relating to the year 2000 compliance was encountered, since data included women born before 1929, and the text files were generated in the two-digit year format. This is a problem inherent in Microsoft products [118] and was overcome by using a macro replacing the year information into four digits prior to its transfer to Excel, and taking into account any reference to years up to and including 1929.

The data used in this work was validated before analysis. Given the large size of the data set, this was an exhaustive and time-consuming task. Working with large databases is recognised to have severe implications in any application in terms of processing time,

storage space and computational memory necessary to carry out the computations, amongst other possible problems. This fact also caused restrictions on the range of software that could be used based on its effective data handling capability. Early consideration also needed to be given to the computational implementations to be eventually used, coupled with considerations of convergence of these computational algorithms.

During the process of validation, several errors and anomalies were identified in the original database. This resulted in a series of corrections or exclusions being made which are detailed as follows:

Corrections:

- Unknown *date of birth*. A cohort of women exists with incomplete dates of birth. If the day is missing the beginning of the month is assumed, whereas if the month is missing, then January is assumed as their month of birth. This accounts for 0.87% (1182 of 136228 entries) of the entire data.
- Data entry irregularities in dates. Other irregularities relating to entries of the date of screening (for example an entry of 1982 when it should have been 1992). Those irregularities were detected and corrected.
- Incomplete *Screening End Code*. Detecting women with a date of screening but without a screening end code. This is due to the screening process being incomplete. Twenty eight cases over the 4 episodes were detected, involving 27 different women. These women are included in the present work. In general, these

women are re-invited to the next episode, as usual, if their age is in the qualifying age range for the next invitation.

- Other data entry irregularities. There were 20 cases with a screening date but without a *screening end code*, nor with a *date of first offered appointment* for the episode. These were due to data entry errors and were corrected and subsequently included in the study.
- Irregularities in episode number. There were 50 cases where a mistake in the entry of the *date of offered appointment* on the computational screening system was detected (data that should belong to the following episode was entered in the previous one). Such faults were eliminated and corresponding patients were accounted for with their correct data.

Exclusions:

- Anomalous *Sx* numbers. Due to the transfer process from the NHS Health Authority database into the Unit database some anomalous *Sx* numbers were identified. For example, some women were assigned more than one *Sx* number and the data also included a number of male patients (jointly accounting for 10,372 cases). These cases were detected by the unit validation system and removed accordingly.
- *Screening variations* of more than 3 years. Forty eight women in the second episode were observed to have more than 3 years' *screening variation*, which may have been due to the movement of such women in and out of the area during the time they were eligible for screening. It may also have been attributed to women

opting to come out of the system for a period of time, but who later returned to the programme. These women were assumed to be an irregularity and were not taken into account for the purposes of the analysis.

- Younger women. Invitations to women younger than expected were detected (due to erroneous registrations on the Health Authority database), and 5 cases were found and removed.
- Age out of range. Women whose ages at appointment fell outside the expected age range (49-75 years old). All women whose age was outside this range were checked and those with errors were excluded from the analysis (5 cases).
- Invitation prior to *DOB*. Date of invitations or screenings occurring prior to the *date of birth (DOB)* were detected. Two patients have been found to have this error, probably due to data entry mistakes. Those women were excluded.

As mentioned earlier, the data relevant to this study was originally held on the MUMPS database, and consisted of the *Sx* number, *date of birth*, *date of first offered appointment*, *type of invitation*, *date of screening* and *screening end code*, all of which were stored for each single invitation. Following the initial clean up of the data, new macros were implemented generating variables necessary to carry out the analysis. These include *age band*, *attendance*, *screening variation*, *round length*, *uptake*, and *coverage*, amongst others. The resulting files were exported into SPSS for statistical analysis.

The population data used for the *coverage* calculations was obtained from the West Midlands Cancer Intelligence Unit [119]. This was necessary because the NHS Health Authority database stores only current information on their population in the area, not

historical data. Such information is collected and stored as aggregates on the Cancer Intelligence Unit databases.

Also, in 1994, the way of reporting statistical information related to the data collection changed, as part of several other changes that were carried out in the Breast Screening Programme. These changes included the way that the KC62 was presented, i.e., from a snapshot approach to a cohort approach.

Furthermore, it should be noted that information pertaining to the *first offered appointment* was first used in the preparation of the KC62's in 1996/97.

3.2 Description of the data and variables involved

3.2.1 Data

Following preliminary preparation, data on 137,051 women was analysed. This only represented non-symptomatic women invited for a preventive programme of breast screening. Data analysed started with first offered appointments in 1989 and closed in the year of the fourth screening round on 31st March 2000.

The women invited were able to take part in up to four different episodes, depending on their age on entering the programme. Women older than 64 are not regularly invited through this process but can be screened upon request. Therefore, the age range of the women included extended to 78 years old.

A small number of women (0.08%) attained a 5th episode. This was unexpected as, from 1989 to 2000, triennial invitations should lead to only four episodes. The reason for this extra episode may be attributed to women having moved to another practice within its catchment period. If this new practice has an earlier invitation round than her original one, the individual concerned has the potential to participate in additional episodes within the 10 year period. The opposite effect can also occur with women having the potential to miss out on an episode. This is overcome by regularly running "fail safe" batches designed to identify women who have not been invited within a 3-year interval. The fifth episode has not been taken into account and will be ignored for the purpose of this work.

3.2.2 Variables

The analysis includes variables separated into two different data sets. The first part of the analysis is based on the set measured at patient level. The variables included in this set are *date of birth (DOB)*, *age at first offered appointment* (FOA), *date of FOA*, *type of invitation*, *date of screening*, *screening variation*, *attendance*, and *round-length*. *Round-length* is measured in two different ways, first as a *difference in years between appointments* and, second, as *years since last screening* (British Standard Slippage). The second part of the analysis involves data measured yearly, including the *number of women invited*, *number of women screened*, *uptake*, *coverage*, *yearly coverage* and *invitation rate*.

3.2.3 Formulae for uptake and coverage

Uptake and *coverage* have routinely been used to measure the performance of screening units [8-10]. However, their particular methods of calculation have varied over the years. A unified model of measurement is therefore desirable in order to carry out comparative studies and to be able to forecast a unit's future performance trend.

These parameters are currently defined by the Breast Screening Programme [77, 78, 80] as:

$$Uptake = \frac{WS}{WI}$$

$$Coverage = \frac{WS_{(3y)}}{Pop - NW}$$

where

WS - number of women screened

WI - number of women invited

WS(3y) - number of women screened aged 50-64 (within the last three years)

Pop - population between 50-64 years old

NW - non-eligible women

Since the Breast Screening Programme cycle is over three years, *coverage* is measured over a three-year period (as it is actually measured by the Breast Screening Programme). The corresponding formula uses the three-yearly moving total of the *number of women screened* (based over the last three years) to give the *coverage* for any particular year.

This is because the whole population is only covered over a three year period. The advantage of this formula is that it results in smoothing out any yearly fluctuation to give the overall trend in the *coverage* for the whole population. Its disadvantage, however, is that it does not explain the actual *coverage* for each year and may be confusing when reported out of context. It is therefore desirable to be able to calculate actual *yearly coverage* figures. Such figures could then be smoothed if necessary.

In order to compute the actual *yearly coverage*, the following alternative definition is introduced based on the actual *number of women screened* in any particular year. It is expected that the maximum value for this *yearly coverage* will be 33.3% (which adds up to 100% over three consecutive years).

Two new formulae are thus introduced in this context:

$$Coverage_{(y)} = \frac{WS_{(y)}}{Pop - NW}$$

$$IR = \frac{WI_{(y)}}{Pop - NW}$$

where:

$Coverage_{(y)}$ - yearly coverage

IR - invitation rate

$WS_{(y)}$ - number of women screened aged 50-64 (invited in the year)

$WI_{(y)}$ - number of women invited aged 50-64 (in the year)

Pop - population between 50-64 years old

NW - non eligible women.

3.2.4 Methods

In Chapters 4 and 5 different statistical methods have been used.

The statistical analysis in Chapter 4 was made using the computer package SPSS for Windows version 9.0.0 [96].

Basic descriptive analysis, frequency & cumulative frequencies are used in Section 4.1.

In the correlation analysis of the numerical variables, in Section 4.2, the Pearson correlation coefficient is used with a significance level of 0.05.

The two-paired sample *t*-Student distribution test with 95% confidence level has been used to test for significant difference of the coverage formula proposed in Chapter 4 (specifically Section 4.3.1).

For Chapter 5 (Section 5.1), in order to establish associations and possible predictors for categorical variables, coefficients such as Lambda (λ), Uncertainty, Phi (Φ), Cramers'V and Contingency have been measured. A comprehensive analysis of these coefficients can be found in references [42, 114, 115].

In this analysis, an association (denoted by x) is classified as follows:

$0.00 \leq x \leq 0.09$	virtually no association
$0.10 \leq x \leq 0.19$	very small association
$0.20 \leq x \leq 0.39$	small association
$0.40 \leq x \leq 0.49$	medium association
$0.50 \leq x \leq 0.69$	high association
$0.70 \leq x \leq 1.00$	very high association

Only those with an association of 0.2 or higher have been considered for analysis.

Contingency tables and their row (columns) proportions based on the marginal totals have been used in order to establish the possible prediction of one variable given a value of the other for those pairs of variables with a sufficient association coefficient value.

A description of the methods mentioned above are explained in detail in Appendix G [31,114].

Chapter 4

Descriptive Analysis

4.1 Results of the descriptive analysis

4.1.1 General variables

First, data relating to individuals is analysed below. A population of 137,051 women had received a grand total of 281,415 invitations. Fig. 4.1 shows the episode relationship and number of women available. As this is a sequential process, therefore, only those women with a 1st episode can have a 2nd episode, and so on.

4.1.1.1 Type of invitation

The type of invitation determines whether a woman is invited for the first time into the system or whether she had previously been invited (even if not in the same screening unit). The pattern of invitations by episode can be followed from Table A.I, where, as expected, the percentage of first call type decreases with the episode number. In the first episode 97% of women are invited to screening for the first time, and only 3% had been invited somewhere else prior to the episode.

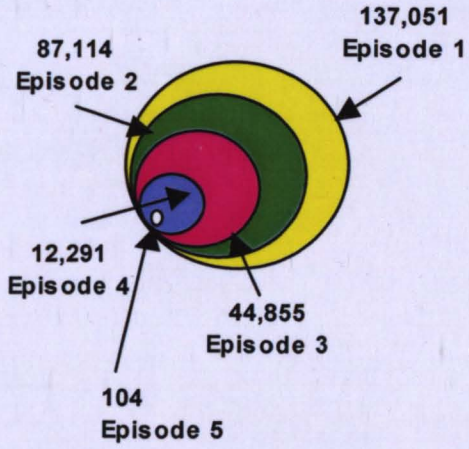


Fig. 4.1 Number of women (numerical breakdown by episode)

"First call" in subsequent episodes (i.e., other than the first) refers to women who were invited in previous episodes but did not attend, and consequently had never been screened. On the other hand, a "Routine recall", in the first episode, refers to women who were previously screened elsewhere. The previous screening history of such women is stored separately on the Unit's database.

4.1.1.2 Age at first offered appointment

Table A.II provides an analysis of the distribution of women's *ages at first offered appointment* for each episode. As expected, the highest percentage of invitations sent out (96.7%) is observed for women between the ages of 50 and 64. Only 0.4% are women aged 65 and older (who may ask voluntarily to be invited but are not included in the rotary system of the screening programme); 2.9% of the total invitations are to women younger than 50 years old, because women are invited by GP practice, not by birth date. Thus for each practice women are called by year of birth, so all women

having their 50th birthday in the current year will be invited, resulting in some 49 year old women receiving an invitation for their first screening [80]. A very small percentage (0.1%) of women younger than 50 are observed to have been invited for their second episode, resulting from re-invitations in periods shorter than 3 years, due to movements of women, GP's or surgeries.

4.1.1.3 Attendance pattern

The first episode was shown to have the lowest percentage of *attendance* at 80.4%, while the fourth episode had the highest at 83%. The general pattern of *attendance* at the screening programme over all the years of its running is that 81% of the invitees attended, with a slight gradual increase in attendance between the 1st and 4th episodes. Table A.III shows that the pattern of attendance was approximately uniform over all episodes.

An analysis of the women who have been invited to four screening episodes led to the following observations [55]. A total of 12,289 women have been invited to four episodes and their *attendance* pattern is shown in Fig. 4.2 and Table A.IV. The best response was in the second episode amongst these women.

While overall 87% of invitations were attended (42,722/49,156), only 71% of women (8,715/12,289) attended all 4 episodes, 93% (11,428/12,289) attended 2 or more episodes and 4% (496/12,289) never attended.

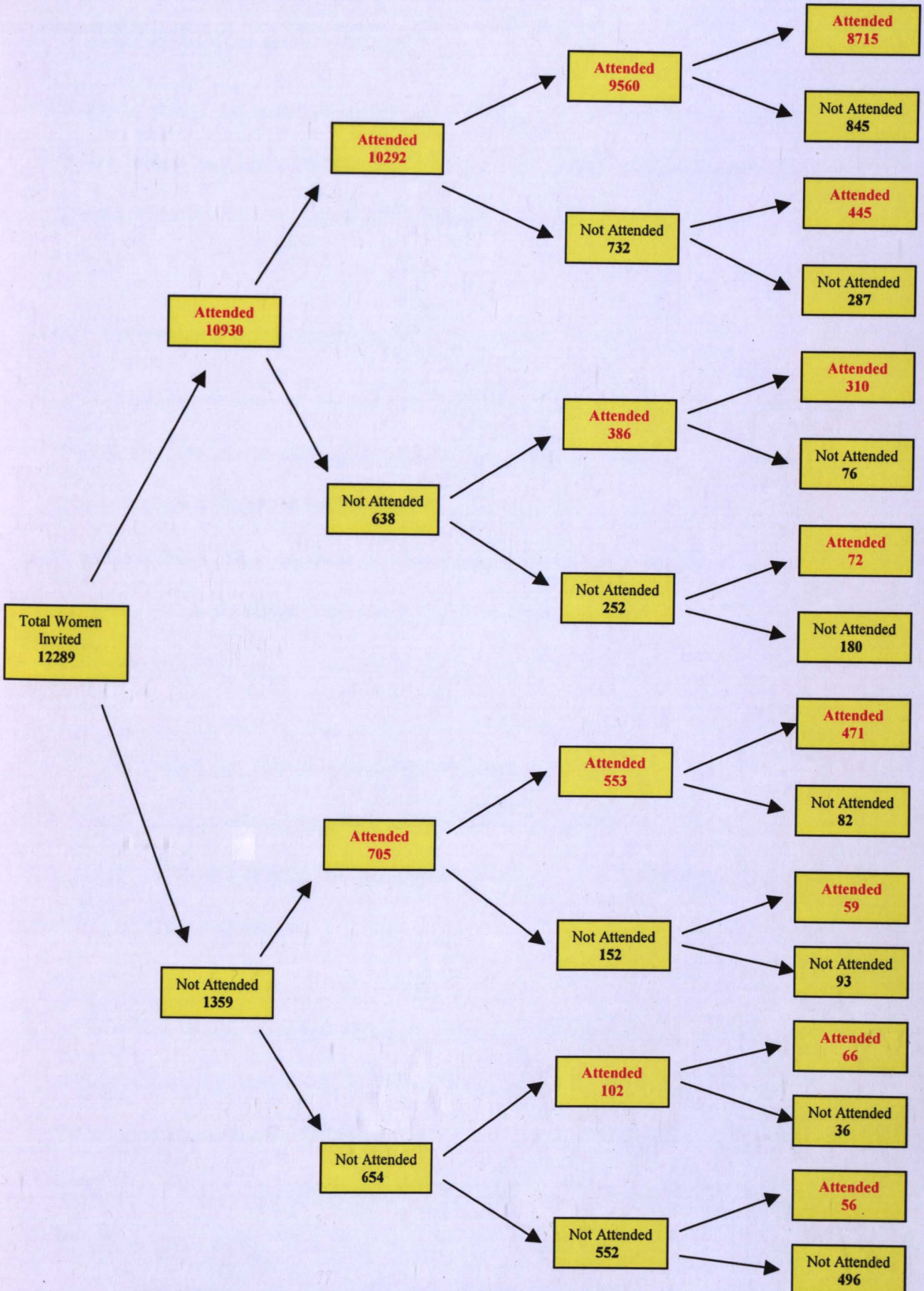


Fig. 4.2 Attendance model for women with four episodes

4.1.2 Attendance performance

As part of the study, the variables in this section were compared in two measuring possibilities, when measured for the calendar year (31st December), against the financial year (31st March). No significant differences amongst them were found.

4.1.2.1 Screening variation pattern for attenders

Screening variation, defined as the time in days between the *date of first offered appointment* and the actual *date of screening*, can result in some women being temporarily excluded from the calculation of *coverage*. If a large percentage of women slip from their initial appointment the *coverage* is at risk of suffering from an artificial reduction. A distribution of the observed *screening variation* by episode is shown in Table A.V.

From 227,736 screenings, 79% were screened on the day of their initial appointment but the percentage of *screening variation* increased with the episode number. Most of this *screening variation* was observed as a change of the date of appointment to a later time than the one initially offered.

However, 99% of women screened attended the actual screening within a period of three months from the date of first offered appointment. Only an infinitesimal percentage of women screened (with respect to the total) had a screening variation out of the screening period of three years, mainly due to movements of those women in and out of the ascribed area.

In general, the pattern of the screening variation from initial offered appointment is shown in Fig. 4.3.

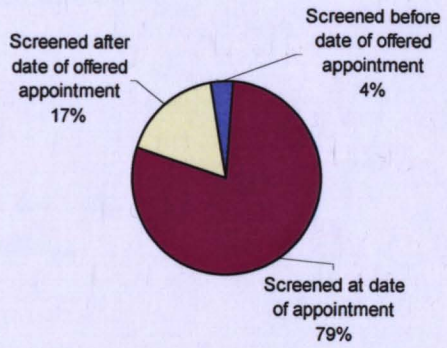
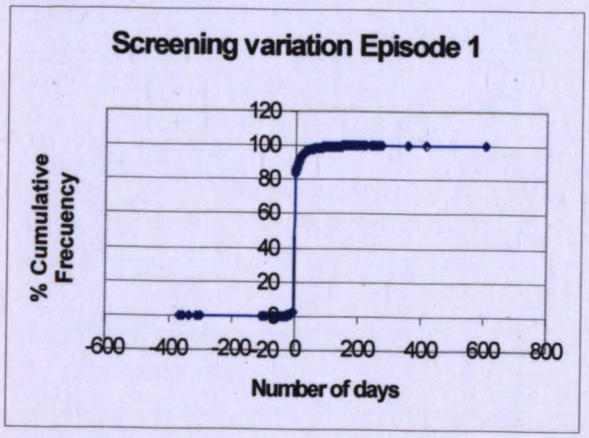


Fig. 4.3 Pattern of the screening variation

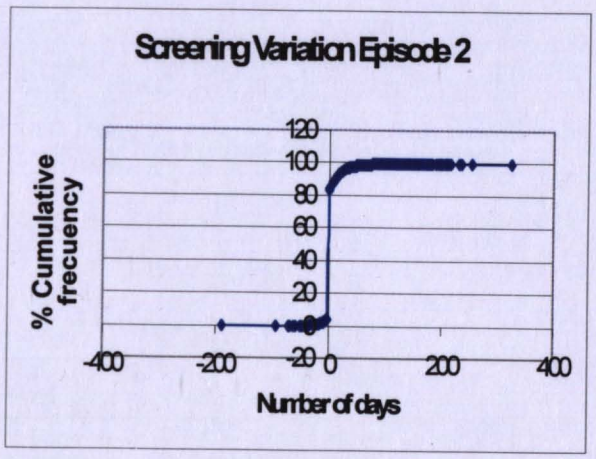
4.1.2.1.1 Screening variation pattern by episode

Given the high percentage of women without *screening variation*, the pattern of *attendance* was analysed graphically divided into three different strands. First, the cumulative frequency percentage is shown (Fig. 4.4 a-d), whereas the second and third sets of graphs show the pre (Fig. 4.5 a-d) and post (Fig. 4.5 e-h) invitation *screening variation.*

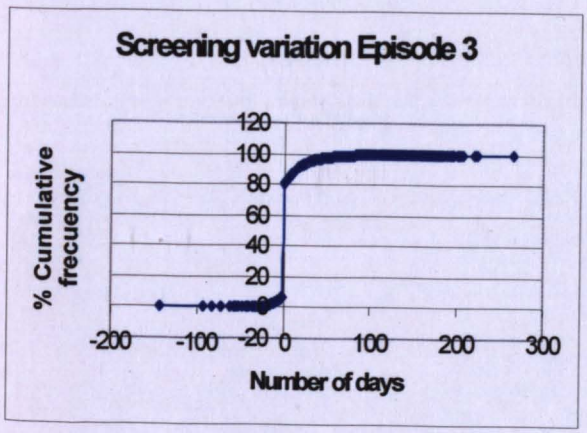
The percentage cumulative frequency charts (Fig.4.4) show a pronounced step change (from 5% to 82%) at the zero value (women who were effectively screened on the date of invitation), because most women attended for screening on the day of invitation. Also, the negative (pre-invitation) *screening variation* branch was less dense and shorter than the positive (post-invitation) one, indicating that women tend to postpone their screening appointments beyond the original invitation date.



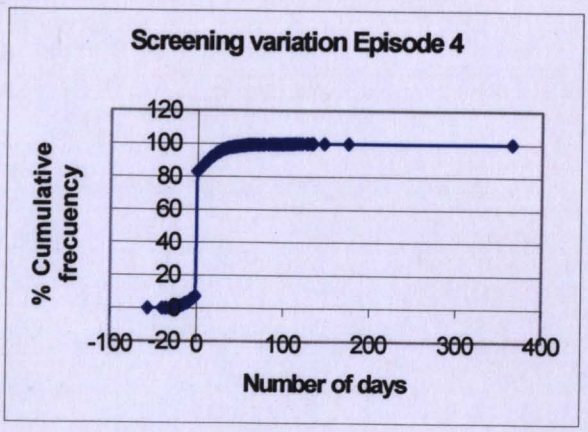
(a)



(b)

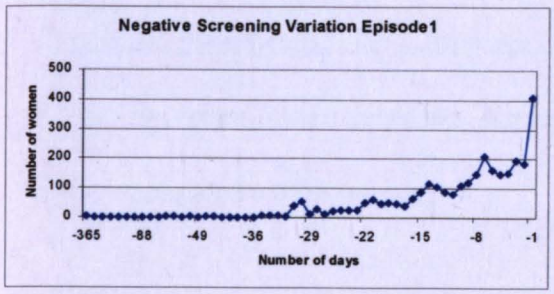


(c)

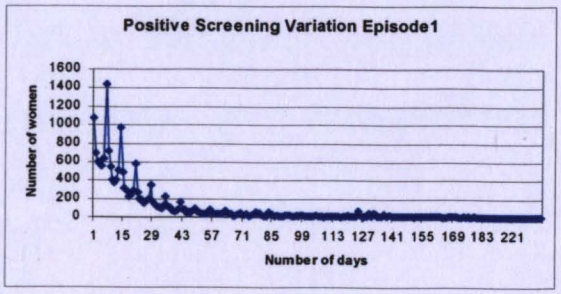


(d)

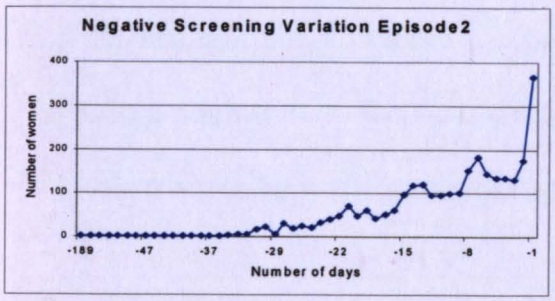
Fig. 4.4 Cumulative frequency of the screening variation by episode



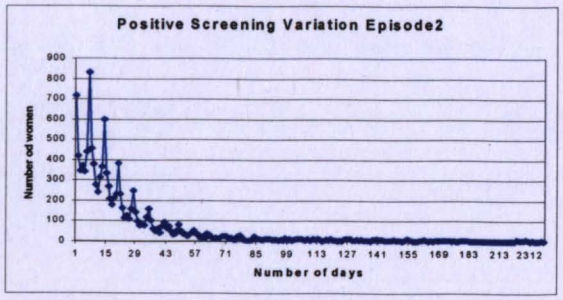
(a)



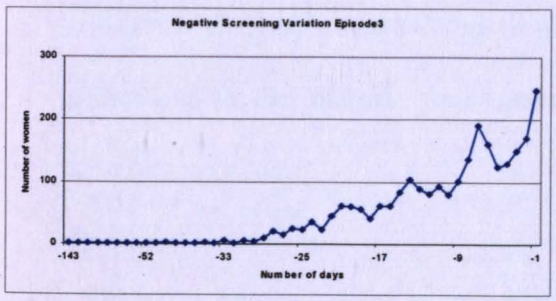
(e)



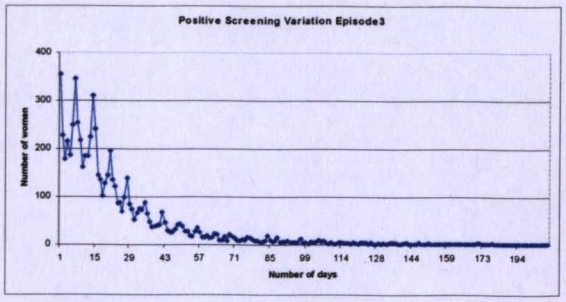
(b)



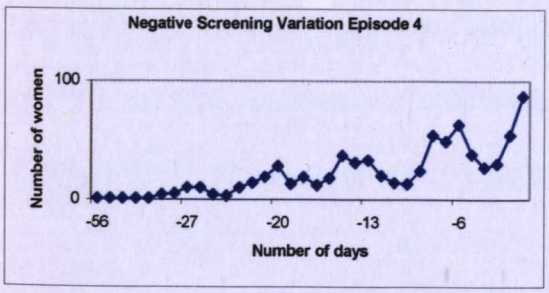
(f)



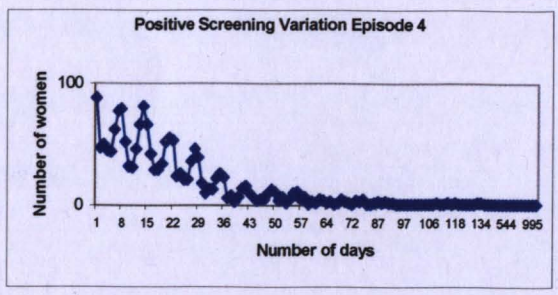
(c)



(g)



(d)



(h)

Fig. 4.5 Negative (pre-invitation) and positive (post-invitation) screening variation by episode. (All attending patients)

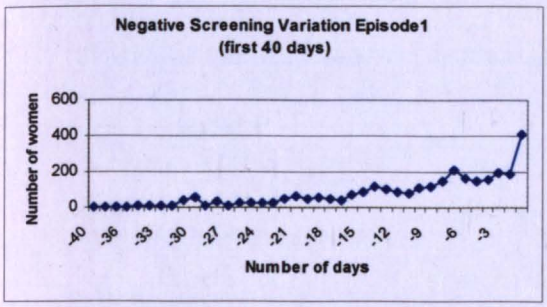
Furthermore, most negative *screening variations* were less than 40 days from invitation; most positive *screening variations* were less than 200 days. Episodes 1 and 2 had less concentrated *screening variations* than episodes 3 and 4, which exhibited a wider change of appointments.

Further analysis of the pre-invitation *screening variation* (Fig.4.5 a-d and Fig. 4.6 a-d), showed that the difference in days between the invitation and the screening decreased slightly with the episode number. The percentage of women with pre-invitation *screening variation* also decreased with the episode.

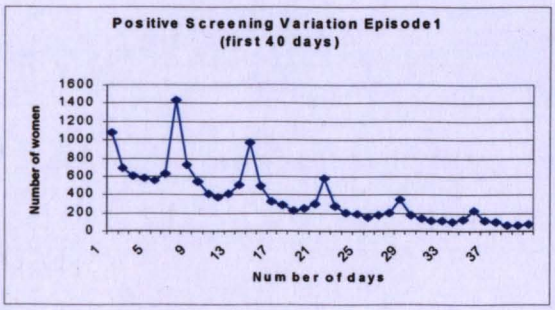
Similarly, an analysis of the post-invitation *screening variation* (Fig.4.5 e-h and Fig. 4.6 e-h) reveals that the percentage of women with post-invitation *screening variation* decreased with the episode number, with the second episode having the longest *screening variation.*

For pre-invitation *screening variation*, the variation for the majority of women was about 30 days from the initial invitation, because the majority of invitations were sent out one month in advance. Conversely, post-invitation *screening variation* varied from about 40 days from initial invitation (episode 1) to almost 70 days (episode 3).

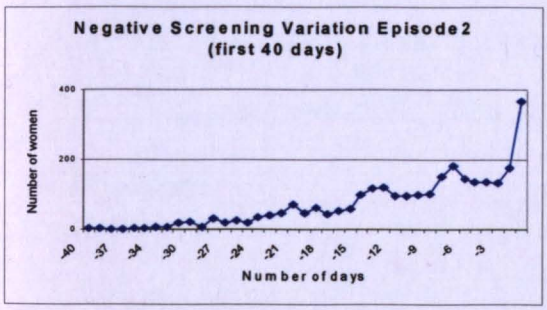
In general, the percentage of women with *screening variation* decreased with the episode number. Since the first 40 days before and after the offered appointment contain the majority of women with changes of appointments, therefore a closer look at this pattern may be of assistance. Fig. 4.6 shows detailed plots of the first 40 days before and after the day of *first offered appointment*. Note the effect of weekends on



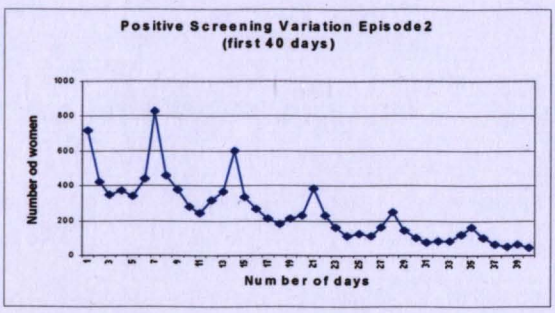
(a)



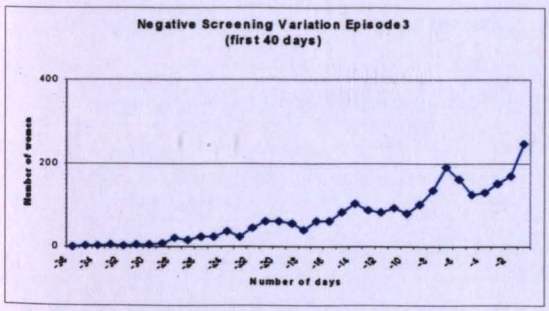
(e)



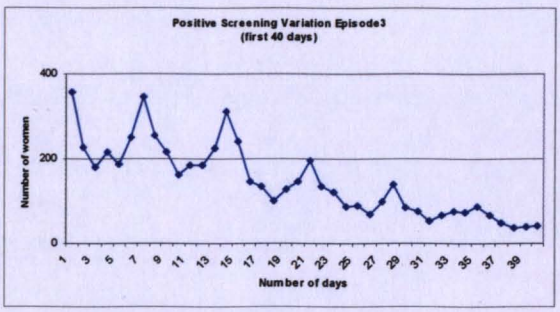
(b)



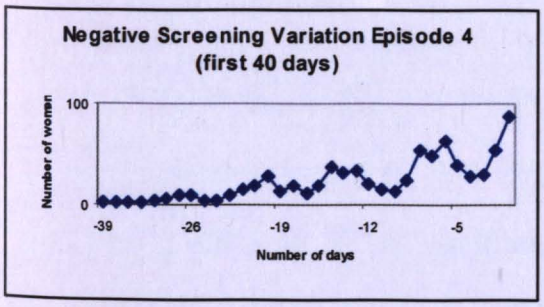
(f)



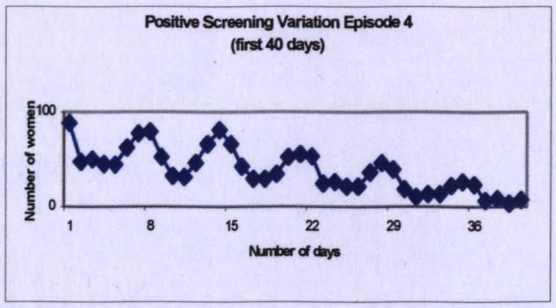
(c)



(g)



(d)



(h)

Fig. 4.6 First 40 days of the screening variation (pre and post–invitation) by episode

the *screening variation* pattern, when there was no invitation. This figure also highlights the differences between pre- and post-invitation *screening variations* for each episode.

4.1.2.2 Years since last screening

This part of the study concerns women who have had two or more episodes. In terms of consecutive invitations, there are 144,364 such invitations involving 123,367 screenings.

Slippage (or *round length British Standard*), is defined as the difference in years between the *date of last screening* and the date of the next *first offered appointment* [79]. Table A.VI gives an analysis of this parameter (however, the term episode is not year-related but invitation-related for each woman). A corresponding value was entered only if a woman had attended the last screening and was re-invited for the next episode.

From a total of 123,367 screenings, 4.7% have a difference of less than three years, 3.5% have a difference of more than three years, and 91.8% have 3 years difference between the date of last screening and the following date of next *first offered appointment*. This result was a measure of the performance of the Unit and the small percentage of *round length* variation was due to women being invited by the GP rotary system and not by *DOB*. Another cause might be due to *screening variation.*

4.1.2.3 Difference in years between appointments

The rotary system for appointments is based on a difference of three years since the last appointment and not on three years difference since the last screening, as the *British Standard Slippage* (BS) for *round length* suggests. This brings into consideration whether the actual system length (*years between appointments*) followed the same distribution as the *round length* (BS Slippage). The variable "*difference between appointments*" is introduced to answer this question.

The *difference between appointments* (also known as *round length*) is the difference in years between the *date of first appointment* of one screening episode and the following one.

Table A.VII gives the results of analysing the difference in years between appointments, taking into account that it is measured only when there are two consecutive invitations.

As mentioned above, 144,364 such invitations were available for analysis. Of these, 5% represented invitations in less than three years, 4.1% in more than three years, and 90.9% were invitations at three years from the previous appointment.

4.1.3 Assessment of the unit

Some of the variables involved in the analysis of performance and assessment of the unit, such as the *number of women invited*, *number of women screened*, *uptake* and *coverage*, are now considered.

The analysis will be divided into three subsections. First variables related to the population are analysed. These variables are measured as aggregates of three overlapping years. No statistical analysis will be performed on this data due to the lack of independence resulting from data overlap. In the second and third part, two new variables (*invitation rate* and *yearly coverage*) will be introduced and discussed.

4.1.3.1 Coverage

Table A.VIII shows the distributions for the variables collected annually (based on the previous 3 years). The incidence of more women than the eligible population being invited for screening should be noted. This is referred to as "over-invitation", and may be due to younger and older women being also invited (and screened), in addition to women in the eligible population range aged between 50 – 64. In addition, the eligible population is understood to be measured at the end of the period and not as an average of the population over the period.

This could either involve missing women who had been in a specified area but who subsequently moved out, or counting women who were newly attached to an area (following their surgery's invitation). The total population (eligible and otherwise) does not only include women who are registered with the NHS, but also those who are outside the NHS system and who are therefore not part of the programme. Thus, an annual approach may solve some of the difficulties raised above, but not all of them.

Fig. 4.7 shows that the *coverage* trend (measured triennially but reported annually), has increased slightly since 1995.

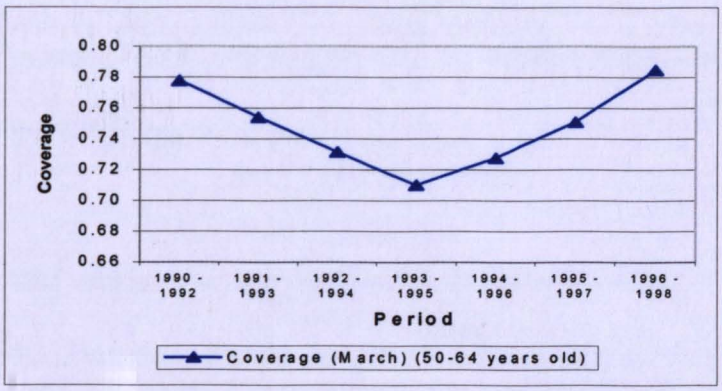


Fig. 4.7 Coverage (3 yearly aggregate)

4.1.3.2 Annual variables analysis (uptake & yearly coverage)

The above variables were analysed at an annual level, and Table A.IX gives the results. However, some of the data are inaccurate. One such inaccuracy was the approximation of the eligible population, since data for yearly coverage included (older and younger) women outside the age range 50-64.

Figs. 4.8 and 4.9 show the yearly patterns of *uptake* and *coverage*.

The graph trend depicted in Fig. 4.8 demonstrates that *uptake* decreased since 1993 and stabilised from 1995. However, for the period 1989-1991, the number of *eligible population* had been decreasing, and increased from 1993 onwards.
With respect to *coverage*, data for years 1989, 1990 and 1999 have not been considered since they were not complete.

As previously stated, given the way in which the screening programme is structured, in order to cover the full population, one third of it should be covered each year. In other words, 33% of the *eligible population* should be screened in one year. The results from the trends analysed suggest that all measurements followed a pattern demonstrating the 33% constraint.

Fig. 4.9 shows that *yearly coverage* fluctuated very little (between 2%-4%), except for a trough in 1995. However, during the period 1994-95 the Unit was known to be subjected to some operational difficulties which are believed to have affected the screening system. Indeed, during 1995 a screening van broke down, consequently, the *number of women screened* for that year decreased. The effect of this irregularity was carried into the following year (1996), when an over invitation was observed. These irregularities should be taken into account when forecasting similar future trends. Attention should also be drawn to the changes related to the invitation system starting from the previous year (1994). Fig. 4.9 also shows the smoothing effect of the 3 yearly aggregate *coverage* compared with *yearly coverage*.

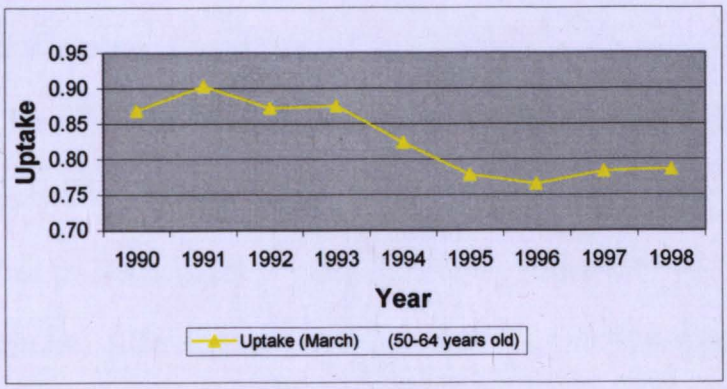


Fig. 4.8 Uptake annual trend

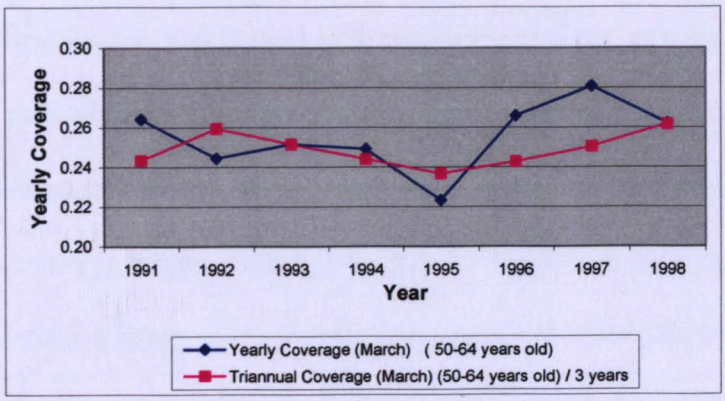


Fig. 4.9 Yearly coverage annual trend

In 1998, more invitations were issued than eligible population as a result of the programme introducing measures to bring round length closer to 3 years. (Table A.VIII)

4.1.3.3 Comparison with the theoretical variables (invitation rate)

As defined previously, the *invitation rate* is the proportion of women invited to the eligible population. Fig. 4.10 shows the *invitation rate* trend (*coverage* that could have been achieved had all women invited been screened), and the *coverage*.

The over-invitation phenomenon referred to in section 4.1.3.1 can also be seen in this figure where, for 1996 and 1997, the *invitation rate* took values higher than 34% of the *eligible population*. In other words, the cumulative *yearly coverage* for the years 1996-98 resulted in 104% of the *eligible population* being covered. This finding can be interpreted in two different (but possibly overlapping) ways. Either the system was reducing the *round length* (inviting women in less than 3 yearly periods) and/or it was including women who were not strictly eligible. The first interpretation is preferred following the knowledge that it is a deliberate policy of the system to reduce *round length*. From the *round length* results previously discussed, between 8% and 9% of invitations, the women were re-invited in a period of less than three years. Interpreting this result in *coverage* terms, roughly, every year 3% of the *eligible population* was invited in less than a three years period, and apparently counted in the formula for *coverage*. Further analysis of *coverage* as currently calculated was carried out and led to suggesting the formula proposed in section 4.3.1 below.

Fig. 4.10 shows a comparison between the *coverage* and *invitation rate* trends.

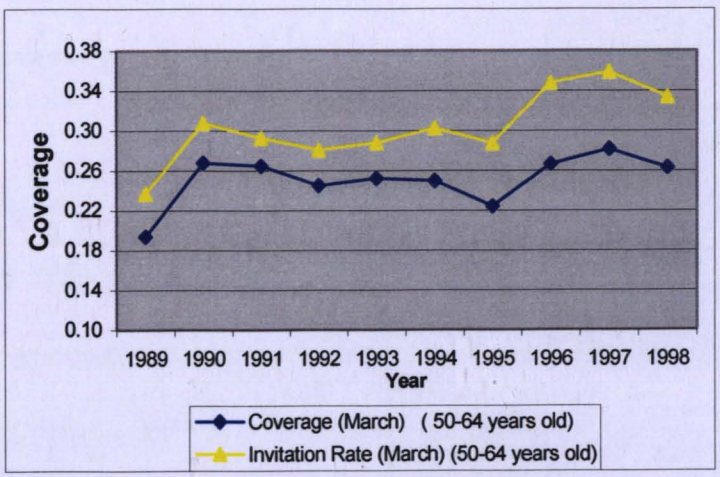


Fig. 4.10 Yearly coverage trend comparison (50-64 years old)

4.2 Other relationships of interest involving quantitative variables

This section is concerned with the analysis of some of the quantitative variables involved. Only those measured at an individual level will be analysed and focusing on aspects such as:

(a) patterns of *screening variation* between episodes. Of particular interest is the recognition that *screening variation* in previous episodes can increase the likelihood of *screening variation* in future episodes;

(b) the pattern of *screening variation* may depend on the age of the patients at invitation;

(c) the *difference between appointments* and the *number of years since last screening* may have similar patterns;

(d) the possible influence of the *screening variation* on the *years between appointments*, i.e., whether the *screening variation* affects the 3-year period round stabilised for the Breast Screening Programme.

The following results were noted [56]:

There was no evidence that *screening variation* in previous episodes increased the likelihood of *screening variation* in following episodes if measured in days. Nevertheless, by grouping the *screening variations* as intervals, the contingency tables results showed dependence with previous *screening variations*. In other words, *screening variation* in one episode was associated with *screening variation* in the following, but

this association was not very pronounced. Nevertheless, if an episode was not subjected to *screening variation*, then this implied a probability between 76% - 80% that the following episode would equally not be subjected to *screening variation*.

A significant dependence was also obtained between the *screening variation* and the *age group of the women at invitation*. In general, the older the women, the higher was the probability of not undergoing *screening variation*.

For the second episode only, the *screening variation* influenced the *round length* in the third episode (*years between appointments* for the second and third episodes). In general, for those women who did not undergo *screening variation*, there was 91% - 95% probability that their *round length* would be 3 years. Only extreme bands of the *screening variation* affected *round length*.

There was also a marked relationship between the *years since last screening* and the *years between appointments* for each specific episode, i.e., for a given episode, the greater the difference in *years between appointments*, the greater was the number of *years since last screening*, (the opposite also applied). This was a measure of the performance of the unit. As noted previously, the appointment system was based on date of appointment and not on date of screening as the programme assumed. Thus, both measurements of *round length* should be related. Also, the variable *years between appointments* contained the variable *years since last screening*. Furthermore, in the hypothetical case when 100% of *attendance* and 0% of *screening variation* is obtained, both variables coincide. This similarity was concluded to be due to the observation that 99% of women slip within a maximum interval of ±3 months from the initial offered

appointment. This conclusion was also strengthened by 98% of the invitations being screened in the first forty days after the initial date of invitation.

4.3 Suggestions for change

Results presented in the previous sections suggest that alternative ways exist for assessing performance in terms of population coverage.

4.3.1 Alternative formula for coverage

In order to avoid the influence of *screening variation* on the calculation of *coverage*, a careful inclusion of women who slip in a range of ±3 months from the date of invitation should be considered. Care, here, refers to ensuring that any overlapping information would be taken into account, i.e., those women invited in a given period should be counted as screened if their attendance at screening fell within the range of ±3 months. Consequently, for the next period, screenings for which invitations fell in the same range should not be counted for that period (since they should have been counted in the previous). A new formula for the calculation of *coverage* for the Warwickshire, Solihull and Coventry Breast Screening Unit can thus be formulated as [55, 60]:

$$Coverage = \frac{WS_{(y)} + WS3M_{(y)}}{Pop - NW}$$

where:

WS(y) - number of women screened aged 50-64 (invited in the year)

WS3M(y) -number of women screened in the following/previous 3 months aged 50-64 (invited in the year)

Pop - population between 50-64 years old

NW - non eligible women.

Since any proposed formula has to be applicable nationally, we have no evidence to support or refute the contention that Warwickshire, Solihull and Coventry is unique in its pattern of *screening variation*. Therefore, we would suggest that *coverage* be reformulated as,

$$Coverage = \frac{WS_{(y)} + WSSE_{(y)}}{Pop - NW}$$

where:

WS(y) - number of women screened aged 50-64 (invited in the year)

WSSE(y) - number of women screened in the following screening extension aged 50-64 (invited in the year)

[*screening extension* is defined as the historical time average within which the desirable percentage of women screened is achieved, taking into account the historical screening variation for a given unit.]

Pop - population between 50-64 years old

NW - non eligible women.

The main advantages of this new formulation is that it does not permit missing or double counting of women, but it allows a unit to take account of women who do not attend on their allocated date of *first offered appointment*.

The main disadvantage is clearly that this method of calculation could vary between different units. In order to have a unified measurement, the desirable percentage needs to be fixed. However, assigning such a percentage falls outside the scope of this study.

The data from the Solihull, Coventry and Warwickshire Screening Unit, has proved that the means of both *coverage* formulae differ significantly using the *t*-Student Distribution means paired samples test assuming 95% confidence (Table 4.I). The mean of the difference is 0.41 and the standard error of the mean is 0.06.

Table 4.I. COVERAGE COMPARISON

Year	Old coverage formula	New coverage formula
1989	0.19	0.2
1990	0.46	0.27
1991	0.73	0.26
1992	0.78	0.25
1993	0.75	0.25
1994	0.73	0.25
1994	0.71	0.22
1996	0.73	0.27
1997	0.75	0.28
1998	0.78	0.26

4.3.2 Reduction of the invitation period

At present, the invitation process is structured on the basis of invitations in a 36-months cycle. This allows for missing information during the calculation of *coverage* due to the *screening variation* effect.

One way of avoiding this loss of information is to maintain the use of the current formula for *coverage* but to reduce the period of invitation, based on the measured *screening variation.* Thus, instead of re-inviting after a period of 36 months, we suggest implementing a re-screening period of

[36 – average screening variation] months.

This recommendation has both advantages and disadvantages. The main advantage is that there is no need to change the current formula for *coverage* and also the recommendation maintains the political message that *slippage* is not acceptable. On the other hand, shortening the period of invitation reduces - but does not avoid completely - the possible effect of double counting due to overlapping periods. Another disadvantage is that it could be subjected to periods of inactivity in terms of screening invitations if there is very little *screening variation* in one particular year (but the average *screening variation* is high) with the consequent cost involved.

The main disadvantage is that screening has to be accelerated across the whole period to undertake 36 months work in, say, 34 months. This could be a serious issue for units already struggling to maintain a 36-months cycle.

4.3.3 Invitation rate

This third recommendation follows from the aim of the Breast Screening Programme being to reduce breast cancer death rates. The best way to reduce breast cancer mortality is accepted to be by detecting the malignancy at a very early stage. In order to achieve this objective women are screened periodically. Therefore, an issue is how effective is the invitation, -as opposed to the screening- process in covering the population, or, how many women in the population receive an invitation.

A new parameter, the *Invitation Rate (IR),* can be defined as:

$$IR = \frac{WI_{(y)}}{Pop - NW}$$

where:

$WI_{(y)}$ - number of women invited aged 50-64 (in the year)

Pop - population between 50-64 years old

NW - non eligible women.

This new rate, measured yearly, should aim to achieve a figure approaching 0.33 every year in order to cover the whole population in the three-years invitation round.

The main advantage of this parameter is that it measures efficiency of the programme in inviting women, and it is not affected by women's non-attendance. To date, the Breast Screening Programme has been mainly concerned with screening efficiency of the

women reached, but it has not independently assessed its capability of reaching women. This shortcoming is covered by the *invitation rate*.

Nevertheless, *invitation rate* does not give an overall picture of the screening programme because it is independent of women's *attendance*. Therefore, it is important not to consider it as a stand-alone parameter, but to introduce it as an additional measurement of the performance.

4.4 Summary of the statistical analysis

A number of results obtained from this part of the study point to the performance and achievements of the Unit. These can be summarised as follows:

- 96.7% of women invited were aged between 50 and 64 years old (*eligible population*), which confirmed the adequate design and functioning of the invitation system with respect to age;
- 81% of women invited attended; however, a decrease of 5-10% in *uptake* since 1993 had been observed;
- 99% of women screened attended screening within a period of three months from the *first offered appointment*. This reflected the satisfactory design of reminders and planning of the screenings;
- 99.9% of women attending completed the screening process. This suggested that the women accepted the satisfactory quality of service;
- 79% of women screened did not undergo *screening variation*; i.e. were screened on the date of the *first offered appointment*;

- Most of the *screening variation* observed was due simply to a change of appointment for a later date from the one initially offered;
- In general, within the first 40 days after the date of invitation, most (98%) of the women with *screening variation* had been screened;
- From those women invited since the first round of running the programme, 93% attended a minimum of half of the invitations, and only 4% never attended;
- Non-attendance at previous episodes did not necessarily imply non-attendance at future invitations;
- 91.8% of screenings had a *round length* (BS) of 3 years, and the similarity with the results for the *difference between appointments* showed that, even if the invitation system for the unit was based on three years from the last offered appointment, it still followed the British Standard for *round length* (three years from last screening);
- *Screening variation* in previous episodes increased the likelihood of *screening variation* in future episodes. However, no undergoing *screening variation* in the previous episode implied a probability between 76% and 80% of not undergoing *screening variation* in the next episode either;
- *Screening variation* and *age group* had been noted to be related. In general, the older the women, the higher was the probability of not undergoing *screening variation*.
- *Screening variation* only affected the *number of years since last screening* for the third appointment. In general, for those women who did not undergo *screening variation*, there was 91% - 95% probability that their *round length* would be 3 years. Only extreme bands of the *screening variation* affected *round length*.
- *Attendance* at a previous episode was a good predictor of *attendance* at the following. The opposite also applied.

- In order to improve *attendance* at the screening episodes, different *age groups* of women should be targeted depending on the episode of invitation they are in.
- The most probable *screening end codes* were S- and S+A-, but the risk of *cancer* by *age group* changed with the episodes. For the 1st and 2nd episodes these include women younger than 50 or older than 64, whereas for the 3rd and 4th episodes those aged 55-64 were included.
- Women with a negative *screening end code* had high probabilities of *attendance* but, for those with a positive end code, the opposite applied.
- The higher the *Townsend deprivation score*, the smaller was the probability of *attendance* by women to invitations. Particularly small was the probability of *attendance* for those women living in areas with a Townsend score 5 or higher (more deprived).
- In general, there is not much difference in the probabilities of *screening variation* occurring independently of the *Townsend deprivation score*.
- No difference in the results of *uptake* and *coverage* was noted whether the screening year ended on 31st of March or on 31st December. The continuation of using 31st March as the date of ending of the screening year is proposed.

Problems in the definition of *coverage* were highlighted. This was followed by a suggestion that a careful inclusion of women who varied their screening date in a range of three months from the date of invitation might be necessary. This inclusion should help to avoid the influence of the less than 3 months *screening variation* in the calculation of *coverage*. A generalisation of this result may be to use the amount of time of the *screening variation* for the particular unit (instead of a fixed 3 months) given a previous agreed percentage of women screened.

Also, in relation to the formulation of *coverage*, in 1996, more invitations were issued than the *eligible population*. The causes of these findings are known and have been discussed in section 4.1.3.2. Counting of women outside the eligible age range had also occurred. A solution may be to include in the *eligible population* the number of women aged 49 years old, but without changing the actual method of invitation. This suggestion is based on the number of women actually screened outside the eligible age due to the invitation scheme.

The significant *coverage* drop in 1995 was known to be due to operational reasons. This had also resulted in a decrease in the number of women screened in 1995. Analysis of this logical impact could be useful in forecasting the effects of future operational hazards.

Finally, from the results obtained during the descriptive analysis, three main recommendations are suggested:

- an improvement in the formula for calculating *coverage*;
- a reduction in the operational invitation period;
- a new performance measure defined as *invitation rate*.

4.5 Analysis of findings and their significance

As stated in the statement of the problem, most studies performed worldwide in relation to this subject are invariably control-based studies. The UK Breast Cancer Screening Programme is a population-based programme, and as a consequence, there is no control group to assist in its assessment. Hence, particular importance is given to proxy measurements acquired and to their efficacy in measuring what is actually ocurring in the programme.

The literature shows that many studies were based on the influence of socio-economic factors on *attendance* and *cancer* mortality. However, very few studies, or even, none have been carried out on the influence of the intrinsic factors of the screening process affecting *attendance* from one episode to the next. Therefore, this statistical analysis has mainly focused on those intrinsic factors involved. Nevertheless, analysis of some general socio-economic factors, such as the *Townsend deprivation score*, has also been performed.

As previously mentioned, *uptake* and *coverage* have been used routinely to measure the quality of screening units. However, their methods of calculation have varied over the years. A unified measurement has been proposed and assessed in this part of the study, which facilitates their comparison, and therefore, permits forecasting future patterns and asses the real impact of the breast screening programme on the reduction of breast cancer death rates.

Another issue is that the rotary system for appointments is based on the difference of three years since last appointment, and not on three years difference since last screening, as the *British Standards Slippage* (BS) for *round length* suggests. This raises the question of whether the real system length (*years between appointments*) follows the same distribution as the *round length* (*BS Slippage*). In order to address this issue, the variable "*difference in years between appointments*" has been introduced. The statistical analysis has proved that, even if the programme is designed to have the recall system of three years from the initial date of offered appointment, the system is still compatible with the British Standard.

Very little work has been conducted on the analysis of the influence that changing screening appointments has on the screening system in general, and on the overall measurements of *uptake* and *coverage* over a given time period, in particular. A new factor influencing the screening programme was identified and named "*screening variation*". The *screening variation*, defined as the time in days between the date of first offered appointment and the actual date of screening, can play an important role in the results for *coverage*. If a large percentage of women slip from their initial appointment within a wide range, they run the risk of not being counted for the purpose of *coverage* for the year analysed. The consequence is that this yields misleading information in relation to the accuracy of the system.

Two important factors to take into account in order to achieve the aim of reducing mortality from breast cancer in the screened population are i) the effective *attendance* of women in response to the screening invitation and ii) the reduction in the range of *screening variation*. *Screening variations* measured in days in previous episodes did

not appear to show an increase in the likelihood of *screening variation* in following episodes. Nevertheless, by grouping the screening variations as intervals, results showed dependence with previous screening variations. A significant dependence was also obtained between the screening variation and the age group of the women at invitation.

Ultimately, the *attendance* of women to screening is a crucial way forward to increase the positive outcome of the Screening Programme. Prediction is possible for *attendance* at one episode given the *attendance* pattern at the previous one. In particular, if women attended in response to the invitations in the previous episode, they are more likely to attend at the next one as well. However, if they did not attend at the previous episode, then there is a high probability that they will not respond to the invitations at the following episode either. However, non-attendance at previous episodes does not necessarily imply non-attendance to future invitations. The results indicate that, in order to improve *attendance* to screening episodes, different *age groups* of women should be targeted depending on the episode of invitation they are in.

Possibilities of prediction of the *screening end code* of a particular woman, given her *age group*, have been explored. This result could play an important role in helping to target those women in high risk *age groups* in order to maximise their *attendance* for screening.

In general, independently of the episode, if the *screening end code* for the previous episode is negative, there is a high probability that the women will attend the

following episode, but if the *screening end code* is positive, then the probability is high for non-attendance.

The influence of socio-economic factors on the women's *attendance* to the screening invitation was measured using the *Townsend deprivation score*. The results showed that the higher the *Townsend deprivation score*, the smaller the probabilities of *attendance* of women to invitations, particularly for those women living in areas with Townsend score 5 or higher (more deprived). This finding suggests targeting those women living in the poorest areas in order to increase their *attendance* for screening.

In order to achieve the objective of detecting the malignancies at a very early stage, women are screened periodically. Therefore, an important issue is how effective this invitation - as opposed to the screening process - is in covering the population. To date, the Breast Screening Programme has been mainly concerned with screening efficiency of the women reached, but it has not independently assessed its capability of reaching all women. The parameter *invitation rate*, formulated and proposed in the present study, addresses this discrepancy.

Other factors playing an important role in the evaluation of the quality assurance of the programme are i) methods of data collection, ii) implementation of data storage methodologies and iii) the derived variables stored in order to facilitate the evaluation. In the present work, a successful transfer of the data from a MUMPS database into a more general and user friendly Microsoft-compatible format was achieved, which facilitated the manipulation of the data for research studies.

Statistical validation of the data has been performed, together with the generation of new variables, as an initial stage of the descriptive analysis. Recommendations have also been made for handling the data following the highlighting of errors in its transfer from the Health Authority database into the Breast Screening Unit database.

Following the availability of the data for analysis, a comprehensive descriptive analysis of all the parameters influencing the screening process stored in the original database was carried out for the first time.

A new analytical approach to the data exploration, investigating the screening episodes from the perspective of the screening history of each individual woman, as opposed to the 3-yearly cycle rounds from the GP perspective imposed by the screening programme, has been introduced. This approach also facilitates the use of artificial neural networks (ANN) in the prediction of *attendance* and *screening variation* of a woman for a particular screening episode.

Chapter 5

Artificial Intelligence and Predictive Analysis

Following the results of the analysis carried out in section 4.5, and having highlighted the necessity of developing a predictive tool in order to detect those women likely not to attend or change their screening invitations date, the following chapters will focus on the possible development of such tools.

Firstly, efforts are concentrated towards the prediction of *attendance* of any particular woman to the screening invitation. This is followed, as a logical extension to the work, by an attempt to predict *screening variation.* In other words, having predicted women likely to attend, then it is attempted to predict likely changes in dates of appointments for these women.

5.1 Relationship and predictors of attendance and screening variation

Important factors to take into account in order to achieve the aim of reducing mortality from breast cancer in the screened population, are the *attendance* of the women in response to the screening invitation and the reduction in their range of *screening variation* if any.

This section investigates the relationships between the different factors affecting *attendance* and *screening variation* in the screening programme, as well as their possible predictors.

Variables analysed in this part include the *age band* of women at invitation, *postal area* in which they live, *Townsend deprivation score* [94] for that postal area, their *attendance* to the invitation, their *screening variation*, the *screening end code* resulting from the screening episode and the *round length* of the screening episode. Independence between pairs of variables has been analysed using contingency tables.

In order to establish associations and possible predictors for categorical variables, parameters such as Lambda (λ), Uncertainty, Phi (Φ), Cramers'V and Contingency coefficients have been measured using 95% confidence level. A comprehensive analysis of these coefficients can be found in [114, 115].

In this analysis, an association (denoted by x) is classified as follows:

$0.00 \leq x \leq 0.09$	virtually no association
$0.10 \leq x \leq 0.19$	very small association
$0.20 \leq x \leq 0.39$	small association
$0.40 \leq x \leq 0.49$	medium association
$0.50 \leq x \leq 0.69$	high association
$0.70 \leq x \leq 1.00$	very high association

Only those coefficients with an association greater than or equal to 0.2 are considered for analysis. These are shown in Table 5.I.

Table 5.I RELEVANT ASSOCIATIONS BETWEEN PAIRS OF VARIABLES

Pairs of variables	Coefficient				
	λ	Uncertainty	Φ	Cramers' V	Contingency
Ageband2 vs. Attendance2	0.75	0.68	1.00	0.70	0.70
Ageband3 vs. Attendance3	0.82	0.80	1.00	0.71	0.71
Ageband4 vs. Attendance4	0.79	0.86	1.00	0.71	0.71
Ageband2 vs. Screening End Code 2	0.73	0.49	0.78	0.35	0.61
Ageband3 vs. Screening End Code 3	0.75	0.68	0.87	0.39	0.66
Ageband4 vs. Screening End Code 4	0.67	0.77	0.86	0.39	0.66
Screening variation2 vs. Ageband2	0.00	0.00	0.81	0.41	0.63
Screening variation1 vs. Round length1	0.00	0.00	0.59	0.16	0.51
Screening variation2 vs. Round length2	0.01	0.03	1.00	0.33	0.71
Screening variation4 vs. Round length4	0.14	0.13	0.51	0.23	0.45
Screening End Code 1 vs. Attendance2	0.01	0.01	0.41	0.29	0.38
Screening End Code 2 vs. Attendance3	0.44	0.21	0.54	0.38	0.47
Screening End Code 3 vs. Attendance4	0.23	0.24	0.44	0.31	0.40
Attendance 1 vs. 2	0.14	0.11	0.41	0.41	0.38
Attendance 2 vs. 3	0.28	0.29	0.63	0.44	0.53
Attendance 3 vs. 4	0.24	0.33	0.49	0.35	0.44
Postal area vs. Screening variation1	0.00	0.01	0.20	0.05	0.20
Postal area vs. Screening variation2	0.00	0.01	0.21	0.06	0.20
Postal area vs. Screening variation3	0.00	0.01	0.20	0.05	0.20
Postal area vs. screening variation4	0.00	0.02	0.40	0.10	0.40

Contingency tables and their row (columns) proportions based on the marginal totals are used in order to establish the possible prediction of one variable, given a value of the other, for those pairs of variables with a sufficiently high association coefficient value. The discussion of the results obtained by this analysis follows.

5.1.1 Associations

The *age band* of a woman at invitation was found to be independent from the determination of either a positive malignant result or a false positive outcome.

All other pairs of variables have some levels of dependency and association.

Particularly high associations were found to exist between the *age bands* of the women, their *attendance* at an episode and the respective *screening end code* (except for the first episode where no association was observed).

In addition, a high association was observed between the *age band* and the *screening variation*, particularly for the 2nd episode.

As expected, *screening variation* has a considerable impact on the *round length* of the screening episode. Furthermore, *screening variation* in a previous episode is associated with *screening variation* in the following, but this association is not very strong.

On the other hand, *attendance* at previous episodes is associated with *attendance* at future invitations.

A good level of association (in the region of 0.4) was observed between the *screening end code* of the previous episode and *attendance* at the following. There is also association with the *screening variation* at the following episode but this association was not pronounced.

The *postal area* in which the women live at invitation is associated (0.2 – 0.4) with the occurrence of *screening variation* but the association with the *attendance* is very weak (less than 0.19).

Also, a weak association was noted between the *Townsend deprivation score*, the *attendance*, *screening variation* and *screening end code* of the women for each episode. Figs. 5.1 & 5.2 show diagrammatically the intra- and inter-screening episode associations.

5.1.2 Predictors

Attendance to a given episode may be accurately predicted given the *age band* of the women. In particular, when analysed for each episode, the second episode shows that the highest probabilities of non-attendance are for women younger than 50 (26%) or older than 64 (22%). For the third episode, non-attendance to invitation assumes high probabilities for those women younger than 55 at invitation (34% in the 50-54 *age band* and 100% if younger than 50). On the other hand, in the fourth episode, the non-attendance probability reaches its peak for those women aged between 50-64 at invitation (24%). For the prediction of *attendance* given the *age band*, each episode can be concluded to have its own characteristics and, in order to improve *attendance*, different age groups of women should be targeted taking into account the screening episode to which they are invited.

The *age band* is also an accurate predictor of the *screening end code* (see definition in Table I) for a given episode (except for the first one). Analysing each episode individually, in the second episode, a preponderance exists of only S- (negative screening) or S+A- (positive screening followed by a negative assessment) end codes.

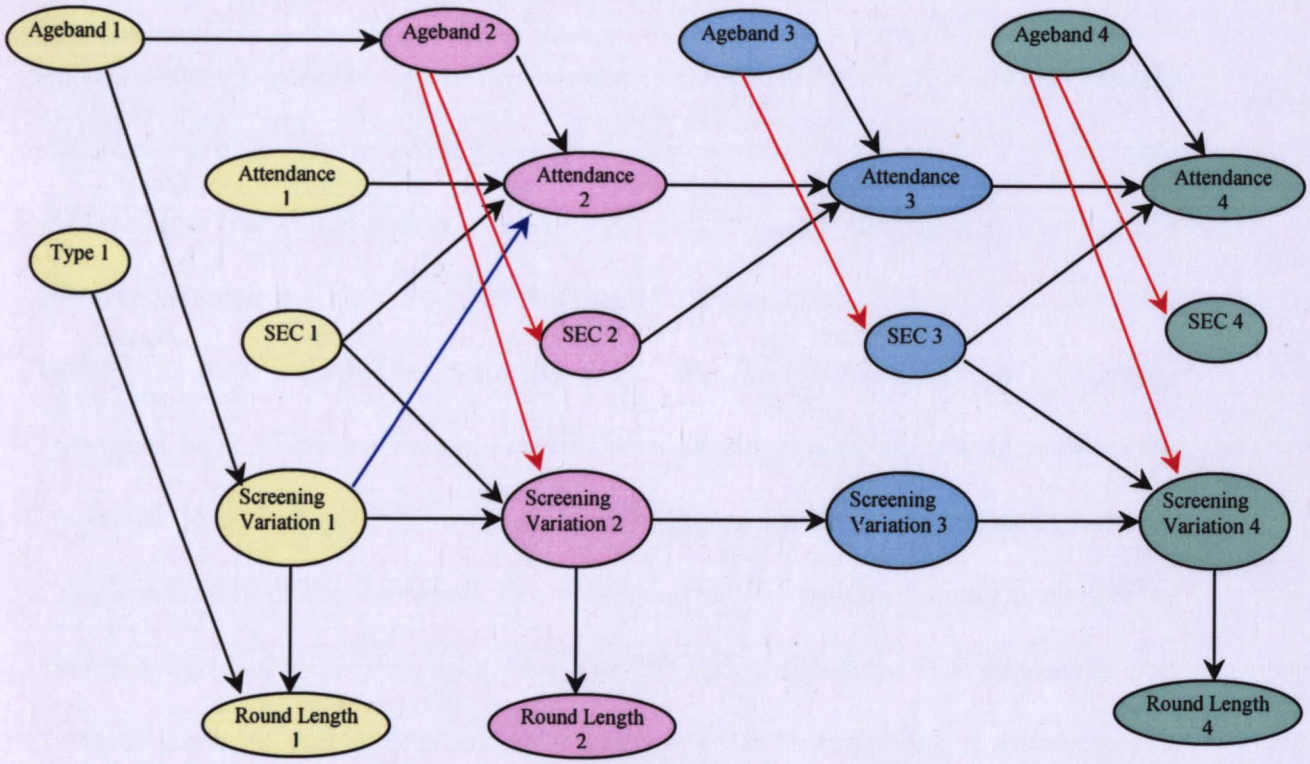


Fig 5.1. Association flow in intra- and inter-screening episode.

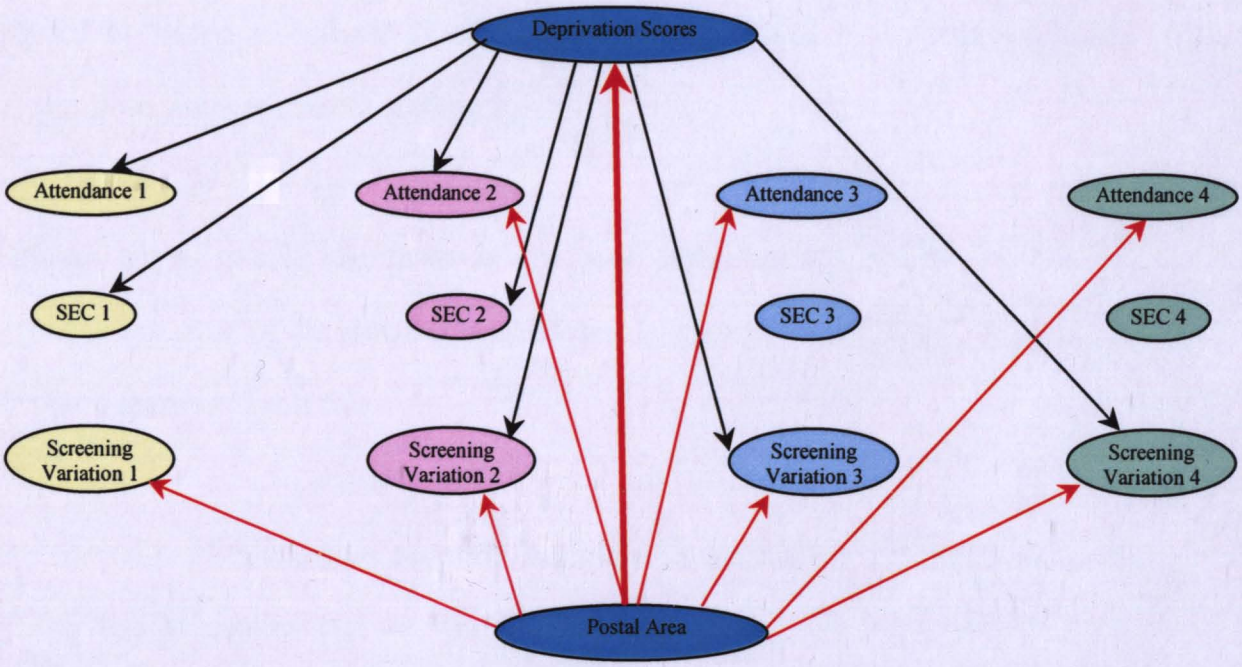


Fig. 5.2. Townsend deprivation scores and Postal area associations by episodes.

Women under 50 years old have the lowest probabilities of an S- result, whilst women with higher probabilities of having a positive cancer result are those who are older than 64 or younger than 50. However, this latter group is more at risk of having a positive or abnormal screening which may not end up in cancer (*false positives*). In the third episode, the most expected outcomes are S- or S+A- independently of the *age group*. Only those aged 55-64 have a very small risk of having a positive cancer outcome for this screening episode. For the fourth episode, the most probable screening outcome is S-, independently of the *age band*. For women aged 55-64 there is a small probability (1%) of having an S+A- outcome. Also, for this *age group* there is a very small risk (<1%) of having a cancer detected. In general, over all the episodes, a *screening end code* of S- is concluded to have the highest probability of occurrence. A result of S+A- frequently occurs in the 2nd and 3rd episodes but decreases considerably in the 4th episode. For the 3rd and 4th episodes the women at a higher risk of having a cancer detected are those aged between 55-64, whereas for the 2nd episode the women at higher risk are those younger than 50 or older than 64.

Although not as equally significant as *age band*, prediction of *attendance* given the *screening end code* of the previous episode can nevertheless be achieved. Analysis by episode is shown in Table 5.II.

As a general conclusion, if the *screening end code* for the previous episode is negative, there is high probability that the women will attend the following episode, but if the *screening end code* is positive, then the probability is high for non-attendance.

Table 5.II

PREDICTION OF ATTENDANCE GIVEN THE SCREENING END CODE (SEC) OF THE PREVIOUS EPISODE

1ST Episode SEC	Probability of Attendance 2nd Episode	2nd Episode SEC	Probability of Attendance 3rd Episode	3rd Episode SEC	Probability of Attendance 4th Episode
No cancer detected	86%	No cancer detected	86%	No cancer detected	92%
Cancer detected	16%	Cancer detected	15%	Cancer detected	9%
S+A*abn* H+	50%	S+A+C*abn*H-	0%	S+A*abn*C+ W+H+	100%
S+A+H+	50%	S+A+C+H-	0%	S+A+H+	50%
S*abn*A*abn*C+ H+	50%	S+A+C-	0%		
		S*abn*A+H+	100%		
		S*abn*A*abn*C+ H+	100%		

Similarly, prediction of *attendance* at an episode given the *attendance* pattern at the previous one is possible. In particular, if the women attend in response to the invitation in the previous episode, they are more likely to attend the next one too. Conversely, if they did not attend the previous episode, then there is a high probability that they will not respond to the invitation to the following episode.

Although relationships were detected between *screening variation* and variables such as *screening end code* and *screening variation* in the previous episode, *Townsend deprivation score* and *postal area* of residence of the women, none of them was identified as an accurate predictor of *screening variation* on its own. This could be due to different factors. Firstly, the structure of relationship between *screening variation* and the other variables might be more complex (i.e. it is not linear), therefore detection is not possible by the prediction coefficients. If this were the case, the use of more complex predictive methods, such as artificial intelligent (AI) algorithms, and in particular artificial neural networks (ANN), may assist in detecting the relationship structure and be able to learn from it. The worst case scenario may be that the information available is not enough for predicting *screening variation*. In such a case, alternative parameters

should be identified. A more detailed study towards this aim is developed in the following chapters.

5.2 Predictive analysis preliminaries

Due to the structure of the invitation process (section 3.1.1) [24], each invitation batch included women who were at different levels of their screening history (i.e. from those in their first invitation up to those in their fourth one).

A challenging feature of the application in computational terms is the existence of expected apparent "missing values". It appears because, even when a woman is invited, she does not always attend the invitation, and/or, even in the case of attendance to the initial invitation, she does not complete her present screening episode. This results in null (but not missed) values for some variables, a problem that is likely to make some predictive algorithms fail or yield poor performance.

In this study, the data set was separated into two initial sets, data with *Townsend deprivation score* information [94] (referred to as the *Townsend data set*), and data without (referred to as the *Postal Area data set*). This approach was opted for based upon the fact that the *Townsend deprivation score* information was not always available for a particular woman (as it was not stored in the Unit's database but in the central NHS Authority database). Previous experiments showed a full separation of the data set for both cases.

In the course of these experiments, it was clearly necessary to select different data structures (and consequently different numbers of women) for the training sets of the predictive models. Thus, the explanation of the data manipulation and variables involved in the predictive models will be separated into two, i) data for the prediction of *attendance*; ii) data for the prediction of *screening variation.*

5.2.1 Data description for the prediction of attendance

Data selection for the development and testing of the models for attendance prediction was subsequently carried out as follows. For each episode and each data set (*Townsend* and *Postal Area*), 20% of women who did not attend, as well as 20% of women who attended screening were randomly selected for the training set. From the remainder of the data, (data that were not previously seen by the models), a further set of 20% randomly selected attending and non-attending of the remaining women was used for validation.

The remaining data (neither used in the training nor validation sets) was used as a test set. However, this data set was the only one that was not balanced in *attendance* terms, i.e., it had more attendance occurrences than non-attendance. Table 5.III shows the resulting data division structure.

Table 5.III DATA DISTRIBUTION BY SET AND EPISODE FOR ATTENDANCE PREDICTION

Episode	Data Set	Training	Validation	Test
1	Townsend	5344	4282	56407
	Postal Area	5372	4300	60523
2	Townsend	3458	2732	36917
	Postal Area	3240	2558	37560
3	Townsend	1716	1352	19647
	Postal Area	1568	1244	19067
4	Townsend	576	458	5673
	Postal Area	646	508	4392

5.2.2 Description of variables involved in the prediction of attendance

Inputs to the models were selected based on the statistical analysis results discussed in sections 5.1.1 and 5.1.2, and depicted in Figs. 5.1 and 5.2 [57, 63].

For the *Townsend data set*, in the first episode, the input parameters were the *postal area number*, *Townsend deprivation score*, *age band* and *type of invitation* for the episode.

The second episode assumes the same inputs as the first, and in addition, included information from the previous episode such as *attendance*, *screening variation* (defined as the difference between the date of invitation and actual screening), *number of tests performed*, *cancer detection* and *false positive occurrence*.

The third and fourth episodes included the same inputs taken into account in the second episode, and had two additional inputs relating to any *previous history of cancer detection* and *previous history of false positive occurrence*.

For the *Postal Area data set*, the inputs were the same as per above, but with the exclusion of the *Townsend deprivation score*.

5.2.3 Methodology for the prediction of attendance

Training of the predictive models was performed under the Clementine SPSS data mining software [121]. The methods studied include Logistic Regression (LR), an Artificial Neural Network Pruning (ANNP) structure, and a Radial Basis Function Network (RBFN).

Assessment of the attendance prediction models performance was carried out using the following parameters [50] defined as:

Sensitivity = (TP/(TP+FN))%

Specificity = (TN/(TN+FP))%

Positive Predictive Value = (TP/(TP+FP))%

Negative Predictive Value = (TN/(TN+FN))%

Accuracy = ((TP+TN)/(TP+TN+FP+FN))%

where:

TP – correct prediction of actual attendance

TN – correct prediction of actual non-attendance

FP – predicted attendance as opposed to actual non-attendance

FN – predicted non-attendance as opposed to actual attendance.

5.2.4 Data and methodology for the prediction of screening variation (SV)

Based on the data used for the prediction of *attendance*, but considering only those women who attended, the following data structure was used for the training and testing of the screening variation predictive models.

In selecting the data to train the models, two different approaches were followed.

The first approach (referred to as *Non-balanced by screening variation classes*) is designed to focus on a data structure appropriate for a model predicting the presence or absence of *screening variation*, without taking into account, at this stage, of the particular screening variation classification.

The whole data set was divided in two (women for whom screening variation was present and those for whom screening variation was absent). The training data set consisted of a random 20% of the subset of those women undergoing screening

variation plus the same number (randomly selected) of those not undergoing screening variation. This guaranteed a 50/50 split of the training dataset. Similarly, the next 20% of data for validation was selected. The remaining data were used for testing the models. Table 5.IV describes the final number of cases per set.

Table 5.IV NUMBER OF RECORDS PER SET NON-BALANCED BY SV CLASSES

Episode	Townsend/Postal Area	Training	Validation	Test
1	Townsend	4224	3398	45170
	Postal Area	4154	3344	49385
2	Townsend	3056	2406	29118
	Postal Area	2900	2312	30144
3	Townsend	2050	1662	14783
	Postal Area	1798	1428	14775
4	Townsend	592	470	4297
	Postal Area	374	312	3368

The second approach (referred to as *Balanced by screening variation classes*) is designed to focus on a data structure appropriate to the prediction of the presence or absence of *screening variation*, where screening variation classification *is* taken into account.

In this approach, data were separated only for training and testing due to the poor availability of cases representing all the possible classifications of *screening variation*. The whole data set was initially separated into groups based on classification of the *screening variation* (including no screening variation). The training data was selected randomly from each group as a proportional (but not equal) representative number of each class. The remaining data were taken for testing the model. Table 5.V gives the number of cases for each data set.

Table 5.V NUMBER OF RECORDS PER SET BALANCED BY SV CLASSES

Episode	Townsend/Postal Area	Training	Test
1	Townsend	1832	50960
	Postal Area	1834	55049
2	Townsend	1823	32757
	Postal Area	1821	33530
3	Townsend	1821	16674
	Postal Area	1757	16244
4	Townsend	677	4682
	Postal Area	535	3519

An extension of the definitions explained in section 5.2.3 is used for the assessment of performance of the screening variation predictive models. These include:

Sensitivity = ((TP)/(TP+TPM+FN))%

(capacity of correctly predicting positive values)

Gross Sensitivity = ((TP+TPM)/(TP+TPM+FN))%

(capacity of correctly predicting positive values with or without correct screening variation classification)

Specificity = (TN/(TN+FP))%

(capacity of correctly predicting negative values)

Positive Predictive Value = ((TP+TPM)/(TP+TPM+FP))%

Negative Predictive Value = (TN/(TN+FN))%

Accuracy = ((TP+TN)/(TP+TN+TPM+FP+FN))%

Gross Accuracy = ((TP+TPM+TN)/(TP+TN+TPM+FP+FN))%

(accuracy of prediction without taking screening variation classes into account)

where:

TP- Correct prediction of screening variation for actual screening variation

TPM - Detection of positive screening variation but with a wrong value of the screening variation class

TN- Correct prediction of actual non-screening variation

FP- Predicted screening variation as opposed to actual non-screening variation

FN- Predicted non-screening variation as opposed to actual screening variation

Attention should be drawn to the fact that the Sensitivity and Accuracy parameters reflect the capacity of the models to predict correctly, not only if the woman undergoes *screening variation*, but also, which degree of *screening variation* she will undergo in the case of a positive prediction. Alternatively, if a simple identification of whether, or not, a woman will undergo *screening variation* (not taking screening variation classification into account), then the parameters to focus on are the Gross Sensitivity and Gross Accuracy, respectively.

Chapter 6

Prediction of Attendance to the Breast Screening Programme

Due to the complexity and variety of predictive methods available for the prediction of *attendance* to the breast cancer screening programme, it has been decided to use a commercial software which was highlighted in the literature review as performing well with medical data. The Clementine environment is not only well recognised for its performance, but also because this environment allows to carry out other data mining tasks and to use the implemented algorithms as libraries in developing multi-method techniques. Although, as with many commercial software, explicit details of the algorithms implemented are not available, a general approach and explanation of the methods can be found in Appendix H.

6.1 Topology of artificial neural network models

A neural network is a simplified model of the way the human brain processes information. It works by simulating a large number of interconnected simple processing units that resemble abstract versions of neurons (Fig. 6.1).

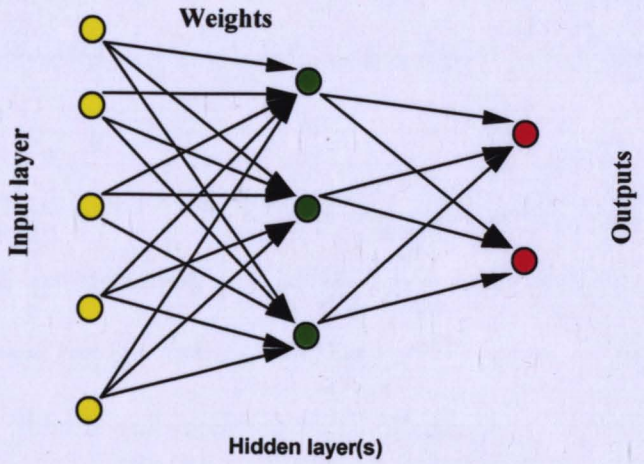


Fig. 6.1 General structure of an artificial neural network

The processing units are arranged in layers. There are typically three parts in a neural network: an input layer with units representing the input fields, one or more hidden layers, and an output layer with a unit or units representing the output field(s). The units are connected with varying connection strengths or weights.

The network learns by examining individual records, generating a prediction for each record, and making adjustments to the weights whenever it makes an incorrect prediction. This process is repeated many times, and the network continues to improve its predictions until one or more of the stopping criteria have been met.

In medical research, the most commonly used ANN is the multilayer perceptron (feed-forward neural networks) with backpropagation training [112]. Those are supervised networks, i.e., there is an output field with observed data that can be used to train a learning algorithm. The algorithm attempts to build a model that produces

predictions that match the observed output values as closely as possible. This external criterion of observed output values is said to supervise the learning process.

As it is well established with ANNs [122, 123], the architecture of a model is directly related to its performance, i.e. performance is dependent on the particular data structure chosen for the training data set and the value of the parameters selected in training the model (its architecture). In the light of this knowledge, several architectures and training parameters have been experimented with for both AI methods (ANNP and RBFN). Results reported here are those of the architectures achieving the best performance obtained.

The use of only one hidden layer yields a more accurate performance in all cases. Thus one hidden layer is adopted throughout.

Table 6.I summarises the number of neurons for the input and hidden layers. The output layer has only one neuron, in all cases, representing attendance or non-attendance.

Table 6.I TOPOLOGY (NUMBER OF NEURONS)

Episode	Data set	ANNP		RBFN	
		Input	Hidden	Input	Hidden
1	Townsend	4	5	4	20
	Postal Area	3	20	3	20
2	Townsend	9	20	9	20
	Postal Area	8	16	8	20
3	Townsend	11	13	11	20
	Postal Area	10	9	10	20
4	Townsend	11	16	11	20
	Postal Area	10	2	10	20

For the ANNP algorithm (refer to Appendix H.2.2) the general parameters used are summarised in Table 6.II:

Table 6.II ANNP GENERAL PARAMETERS

Parameters	Value	Description
Hidden rate	0.15	number of hidden units to be removed in a single hidden unit pruning
Hidden persistence	6	number of hidden unit pruning operations to be performed if no improvement is seen
Input rate	0.15	number of input units to be removed in a single input pruning
Input persistence	4	number of input pruning operations to be performed if no improvement is seen
Persistence	100	number of cycles for which the network will train before attempting to prune if no improvement is seen
Overall Persistence	3	number of times to go through the hidden unit prune / input prune loop if no improvement is seen before training stops
α	0.9	momentum term used in updating the weights during training. It tends to keep the weight changes moving in a consistent direction. Higher values of α can help the network avoid local minima
Initial η	0.3	η is the learning rate which controls the weight adjustment at each update. η changes as training proceeds. Initial η represents its starting value. During training, η start at initial η, decreases to low η, then is reset to high η and then decreases to low η again. These last two steps are repeated until training is completed. η decay specifies the rate at which this parameter decreases as the number of cycles between high and low
η decay	30	
High η	0.1	
Low η	0.01	

For the RBFN algorithm (refer to Appendix H.3) the general parameters used are summarised in Table 6.III:

Table 6.III RBFN GENERAL PARAMETERS

Parameters	Value	Description
Number of RBF clusters	20	number of radial basis functions to use. It corresponds to the size of the hidden layer
Persistence	30	number of cycles for which the network will continue to train if no improvement is seen
α	0.9	momentum term used in updating the weights during training. It tends to keep the weight changes moving in a consistent direction. Higher values of α can help the network avoid local minima
η	computed automatically	η is the learning rate which controls the weight adjustment at each update
RBF overlap	1.0	the hidden units in an RBFN represent radial basis functions that defines clusters, or regions, in the data. This parameter controls the extent of overlap between those regions. Normally during training, records affect only the clusters to which they are closest. By increasing this parameter, the size of the region associated with each hidden unit is increased, allowing records to affect more distant clusters

6.2 Comparing predictive methods results

In developing the relevant predictive models, three main methods are employed. Two of these are artificial intelligence methodologies, namely i) the artificial neural network pruning algorithm (ANNP), and ii) the radial basis function networks (RBFN). For comparison purposes, the traditional logistic regression (LR) method is also used.

Investigation of the results obtained using the training sets (Fig. 6.2 a & b), shows that the three methods (LR, ANNP and RBFN) perform in a very similar fashion for the first episode, with a slight improvement achieved through the AI algorithms (ANNP and RBFN) with respect to the statistical method (LR). Nevertheless, for this episode, LR

perform better in terms of specificity. The highest accuracy values are obtained by the RBFN for the *Postal Area set* (53.9%) and by the ANNP for the *Townsend set* (58.7%).

For the other episodes, the difference between the performance of LR and the AI methods increases significantly by as much as 27% in accuracy in favour of the AI models, for the second and third episodes. The difference between both AI models is small, ranging from 0.37%, (in favour of the RBFN model), obtained in the first episode for the *Postal area data set*, up to 5.9% in the worst case, (in favour of the ANNP model), obtained in the fourth episode for the *Townsend data set*.

Over the four episodes, the ANNP generates more accurate predictions, reaching an accuracy of 77.3% in the second episode for the *Townsend set*.

The accuracy of prediction is shown to reach its peak in episodes 2 and 3. It decreases slightly in the fourth episode, presumably due to the relatively "small" number of data available compared with the other episodes (Table 5.III). The low accuracy results obtained in the first episode were expected due to the lack of information about the women's response to screening prior to participating in the programme.

Results obtained for the validation set are very similar to those obtained for the training data set. Only here, the difference between the two ANN models is less obvious. A more detailed analysis of the results for the validation and test data sets is discussed in section 6.5.

The ANNP produces the highest accuracy (75.7%, obtained in Episode 3 for the *Townsend data set*), sensitivity (94.4%, obtained in Episode 2, *Postal Area set*), and specificity (73.6%, obtained in Episode 4, *Postal Area set*).

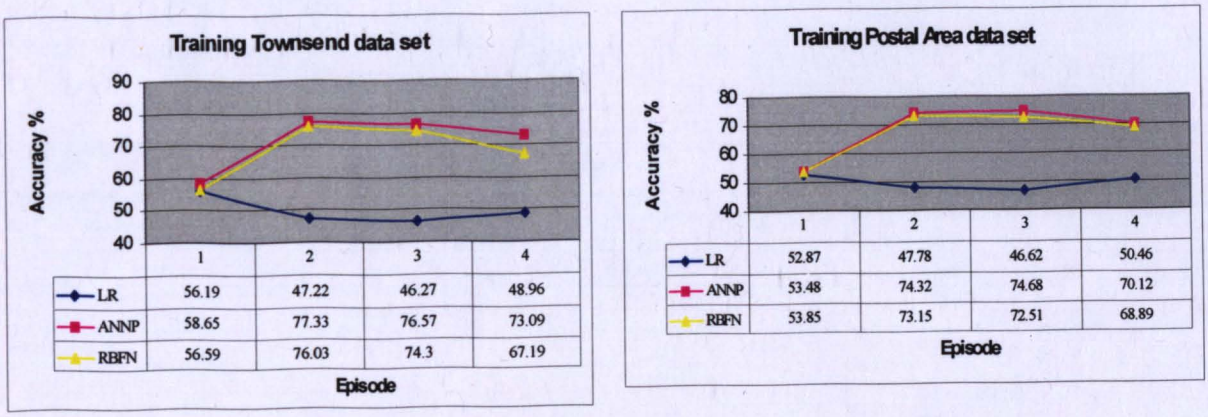


Fig. 6.2 Accuracy of results

The test set reveals similar results to those obtained by the training and validation sets. However, this time the poor performance of the LR is distinctly visible for all episodes. This was expected due to the limited ability of the method in dealing with the "missing data" discussed in previous sections.

Figs. 6.3 and 6.4 show some examples of the performance of the models for the test data set.

From a model performance point of view, the AI models (i.e. ANNP and RBFN) can be said to achieve a better performance in terms of predictive accuracy, sensitivity, specificity and positive and negative predicted values, than the statistical model (LR). This result is expected given the inherent poor management of missing values

involved in this latter method, and the particular data structure of the breast screening invitation process.

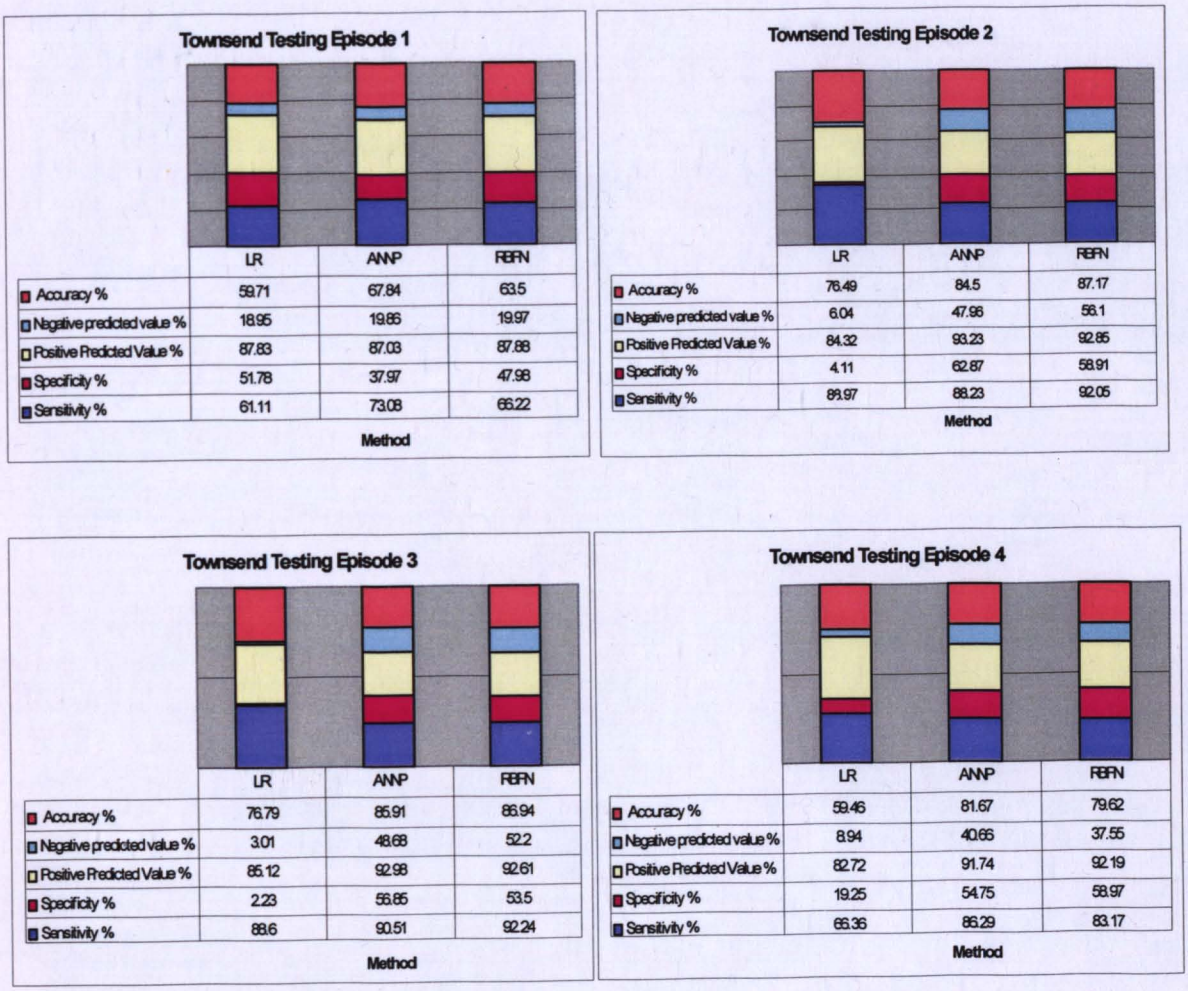


Fig. 6.3. Performance comparison for Townsend test data set (%)

As mentioned above, both ANN models analysed achieve better predictive performance than the statistical model. Moreover, significant differences between the performance of ANNP and RBFN have not been detected which could point to the preference of one method over the other.

Therefore, both AI models are concluded to be suitable for predicting *attendance* to the NHS Breast Cancer Screening Programme for a particular episode. The data from which these conclusions are derived is given in Appendix C (Tables C.I – C.VIII)

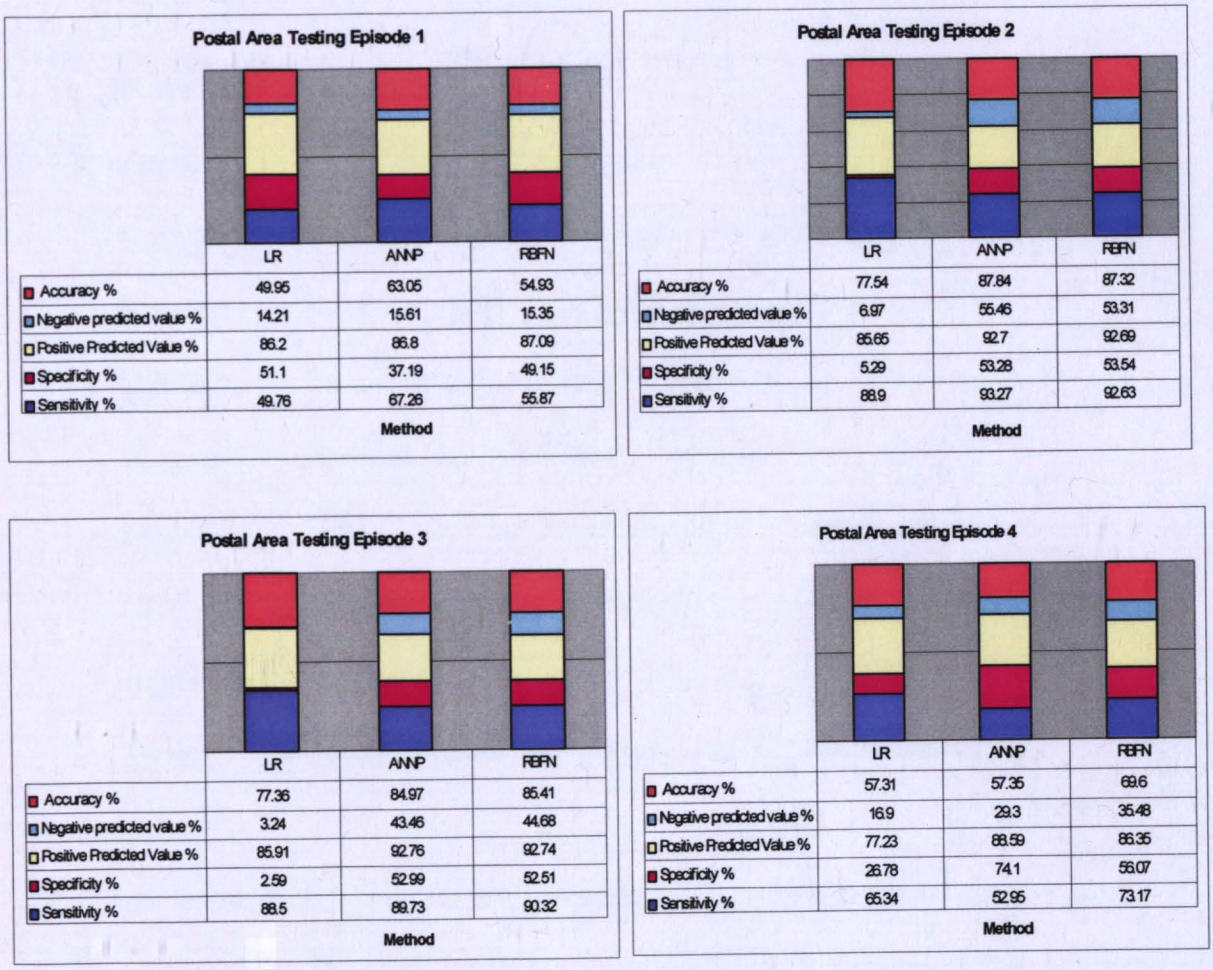


Fig. 6.4. Performance comparison for Postal area test data set (%)

6.3 Predictive feature analysis

The analysis of the predictive features comes as a natural extension of the statistical predictive analysis. The main issue being which features are recognised by the artificial intelligence methods as the main predictors of outcome.

In order to provide an answer to this question, an analysis of the predictive features identified by the three predictive methods has been carried out. Table 6.IV gives a summary of the results. The following abbreviations apply:

Postanum = Postal area number

Townsref = Townsend deprivation score

Screvar = Screening variation in the previous episode

Numtest = Number of tests in the previous screening episode

Cancer = Cancer detection in the previous episode

Histcan = Previous history of cancer detection

Falsep = False positive occurrence in the previous episode

Histfp = Previous history of false positive occurrence

Ageband = Age band

Type = Type of invitation to the episode (first call or recall)

Table 6.IV shows that the most important predictors for all the episodes are: the *postal number of the area of residence* and the *Townsend deprivation score*. Although not applicable to the first episode, important predictors identified for most episodes are: *age band*, *screening variation* in the previous episode, inputs related to cancer detection (*Numtest, Cancer* and *Histcan*), and *false positive occurrence*. Furthermore, as more inputs are available to the models (due to the availability in following episodes of previous screening history), these were selected by all three methods. Hence, it can be assumed that any previous screening history has a direct impact on the prediction of *attendance* in following episodes.

Table 6.IV MAIN PREDICTORS

Episode	Predictors	
	Townsend set	Postal Area set
1	♦ Postanum	♦ Postanum
2	♦ Postanum ♦ Townsref ♦ Screvar ♦ Numtest ♦ Falsep	♦ Postanum ♦ Cancer
3	♦ Postanum ♦ Townsref ♦ Screvar ♦ Cancer ♦ Histcan ♦ Histfp ♦ Falsep ♦ Ageband	♦ Postanum ♦ Screvar ♦ Cancer ♦ Histcan ♦ Falsep ♦ Histfp ♦ Ageband
4	♦ Postanum ♦ Townsref ♦ Screvar ♦ Numtest ♦ Cancer ♦ Histcan ♦ Falsep ♦ Histfp ♦ Ageband ♦ Type	♦ Postanum ♦ Numtest ♦ Falsep ♦ Histfp

Results obtained from this assessment show no significant difference between the analysed methods in the selection of the predictors These predictors were identified by all three models analysed (LR, ANNP and RBFN).

6.4 Decision trees and rule induction analysis of attendance prediction

In an attempt to present an explanation of the operation of ANNs, this section gives an insight into the decision trees and rule induction (Appendix D) analysis on which the AI models previously discussed are based.

Rule induction in the prediction of a particular feature comes as a natural extension of the decision trees; for this study the *attendance*, or not, of a particular woman to her screening invitation is this feature.

The optimum outcome for the screening units in practical terms is the detection of those women who are likely not to attend their screening invitation so that appropriate measures can be taken to encourage their *attendance*. Drawing decision rules that will identify general patterns for those women is a step forward towards this outcome. Applying rule induction techniques to the models generated by the ANN for each episode, whether the *Townsend deprivation score* information is present or not, the following decision trees for attendance, and in particular rules for non-attendance, have been identified. Note that each rule set can be followed from the full decision tree from which it is deduced.

A total of 16 different sets of rules and decision trees have been identified depending on the screening episode a woman is invited at, the availability or not of *Townsend deprivation score* information and the particular AI model used to obtain the prediction.

A summary of the decision rules and trees obtained for each episode is presented in Tables 6.V – 6.VIII.

Table 6.V SUMMARY OF DECISION RULES AND TREES FOR EPISODE 1

Episode	Data set	Method	Decision rules for non-attendance			Decision trees for attendance		Errors of the method
			Number of rules	Discussion	Section ref.	Figure	Discussion	
1	Postal area	ANNP	2	based on the values of two inputs, the type of invitation issued, and the age of the women at invitation	D.1.1	D.1	the age band to which a woman belongs at invitation plays the most important role in determining her attendance	it predicts attendance when no invitation has been issued (Ageband1=0). This is an error due to the initialisation of the variable when handling women at different levels in their screening history, and should not occur in practice as all women are invited to the first episode (otherwise they will not be in the database in the first place)
		RBFN	6	the postal area where the woman lives, her age at invitation and the type of invitation are the main factors covered by these rules. The construction of the rules shows the ability of this method to detect more complicated relationships between the variables	D.1.2	D.2	it bases the prediction of attendance, in the first place, on the postal area where the woman lives, and secondly on the age band to which she belongs at invitation	
	Townsend	ANNP	12	these rules interrelate the woman's postal area, the Townsend deprivation score for the area, her age at invitation, and the type of invitation	D.1.3	D.3	shows that the main predictor on which this model is founded is the Townsend deprivation score information, but the wider ramifications of the tree points to a very complicated relationship between the attendance and the known predictors	assumes attendance of the woman when there is no Townsend information available. It also predicts attendance in some cases where there is no invitation. These theoretical problems should not appear in practice as the data used have Townsend information for all women and all have had a first invitation. The appearance of these errors is due to the initialisation process of the variables before training
		RBFN	2	based mainly on the Townsend deprivation score value	D.1.4	D.4	the social deprivation score and the type of invitation are the main predictors of attendance chosen in this case	predicts attendance when there is no invitation (given by the variables ageband1=0 or typebin1=0)

Table 6.VI SUMMARY OF DECISION RULES AND TREES FOR EPISODE 2

Episode	Data set	Method	Decision rules for non-attendance			Decision trees for attendance		Errors of the method
			Number of rules	Discussion	Section ref.	Figure	Discussion	
2	Postal area	ANNP	4	mainly based on the attendance history of the woman to screening and whether or not breast cancer has been previously diagnosed	D.2.1	D.5	gives higher importance, in terms of prediction of attendance, to the type of invitation and the woman's postal area	prediction of attendance when there is no invitation. This is due to the need to fully instantiate the variable typebin2 (which includes a zero value when there is no invitation). In practice, the algorithm should never take this path as there are no women satisfying this condition
		RBFN	5	simple rules focusing on the attendance of a woman to the previous episode, and the results obtained by the screening process. This method addresses the importance that the occurrence of false positives and the number of tests involved in a single screening episode has in the prediction of non-attendance.	D.2.2	D.6	the main predictor selected in this model is the attendance of the woman to the previous episode invitation. The second predictor of importance is the occurrence or not of false positive results in the previous screening episode.	
	Townsend	ANNP	14	most of these rules are complicated interrelations between the postal area where the woman lives, the Townsend deprivation score for the area, the age band to which the woman belongs at invitation, and screening variation in the previous episode. Non-attendance to previous invitations and diagnosis of cancer in the previous screening constitute simple rules of non-attendance for this model.	D.2.3	D.7	a complicated non-linear relationship of the predictors of attendance is clearly shown in the multiple ramifications of the tree	two theoretical prediction errors for this case, the first one being the prediction of attendance in all the cases where there is no Townsend information available. The second error is the prediction of attendance when there is no invitation to the episode. Both errors do not occur in practice, consequently, no prediction should follow those branches of the tree
		RBFN	6	a complicated interrelationship between factors such as the Townsend deprivation score of the area where the woman lives, her age, and if she undergo screening variation or have had false positive results in her previous episode; are the base for four of the rules in this model . The other two rules, are simple decisions based on the woman's non-attendance to the previous screening episode and if she has been diagnosed with cancer on that occasion.	D.2.4	D.8	based its attendance prediction firstly on the woman's attendance to the previous invitation, and secondly on the detection or non-detection of breast cancer in the previous screening episode. The method in this case, shows a tree with simple decision branches.	

Table 6.VII SUMMARY OF DECISION RULES AND TREES FOR EPISODE 3

Episode	Data set	Method	Decision rules for non-attendance			Decision trees for attendance		Errors of the method
			Number of rules	Discussion	Section ref.	Figure	Discussion	
3	Postal area	ANNP	4	two simple rules related to the woman's non-attendance or her diagnosis of breast cancer in the previous episodes, and two rules interrelating her age at invitation with previous history of false positive screening results or the postal area where she lives at invitation	D.3.1	D.9	a simple branched decision tree with the most important variables, in terms of attendance prediction, being first the attendance at the invitation in the previous episode, and secondly, the detection or non-detection of breast cancer in that episode	predicts attendance when there is no invitation
		RBFN	2	simple rules, one involving the non-attendance of the woman at her previous screening episode, and the second, a combination of the postal area where she resides at invitation and the occurrence of false positive results in the previous screening episode	D.3.2	D.10	a simple decision tree showing the attendance of the woman at her previous screening invitation as the main predictor of attendance to the episode. The second best predictor is the occurrence or not of a false positive result during the previous screening process. Other factors being taken into account are the postal area where the woman lives at invitation and her previous history of breast cancer	no errors were detected
	Townsend	ANNP	2	simple rules ; the first being the woman's non-attendance to the previous episode, and the second her diagnosis of breast cancer in the previous episode	D.3.3	D.11	the attendance of the women at their previous screening invitation is the main predictor of attendance. The other good predictor is the diagnosis or not of breast cancer during the previous screening process. This very simple model identified no other predictors of attendance.	
		RBFN	2	a very simple model based on only two rules. The first one is based on the woman's attendance to her previous screening episode, and the second, on the number of tests involved in her previous screening episode.	D.3.4	D.12	only the attendance at and the number of tests in the screening process for their previous episode are observed to be the predictors of the woman's attendance at the episode	

Table 6.VIII SUMMARY OF DECISION RULES AND TREES FOR EPISODE 4

Episode	Data set	Method	Decision rules for non-attendance			Decision trees for attendance		Errors of the method
			Number of rules	Discussion	Section ref.	Figure	Discussion	
4	Postal area	ANNP	2	the first rule is the non-attendance of the woman to her previous screening episode, and the second is a relationship involving the postal area where the woman lives, her age at invitation and the occurrence or not of false positive results in her previous screening	D.4.1	D.13	the postal area where the woman lives is the main predictor of attendance at the screening episode, followed by her age at invitation and her attendance or not to the previous episode	no errors were detected
		RBFN	3	two of them are simple rules related to the woman attendance to her previous episode, and her diagnosis on it of breast cancer or not. The third rule is a relationship between the postal area where she lives, her age at invitation and the occurrence or not of false positive results in her previous screening episode.	D.4.2	D.14	the attendance of the women to their previous episode is the main predictor. The second best is the postal area where the woman lives at invitation	
	Townsend	ANNP	2	a simple rule focusing in the woman attendance to the previous screening episode, and a second one interrelating several factors including the postal area where the woman lives, the Townsend deprivation score for that area, the possibility of the woman undergoing screening variation in the previous episode, and her age at invitation	D.4.3	D.15	the method bases its predictions on the type of invitation in the first instance, and then on the attendance or not of the woman at the previous screening episode. The branches suggest a non-linear relationship between the factors involved	the model incurs several errors that can be summarised as, first, predicting attendance in cases where there is no invitation; and second giving general prediction in cases when the Townsend deprivation score is not available
		RBFN	3	the two first ones are simple rules involving the attendance of the woman to her previous screening episode, and the diagnosis of cancer in such episode. The third one combines information related to the postal area where the woman lives at invitation and the Townsend deprivation score for the area	D.4.4	D.16	attendance and cancer detection in the previous episode are the most important predictors of the attendance at the episode	no errors were detected

6.4.1 Decision rules and decision trees summary

Each episode and data set are seen to have different decision trees and rules for non-attendance, based on the predictive method used and the availability of social deprivation information for the area where the woman lives at invitation. Nevertheless, some results are common to most generated models. For the first episode, the *type of invitation*, and the *age of the woman* appear to play the main role in taking a decision in favour of the woman's non-attendance to the episode. For all the following episodes, *non-attendance* at the previous screening episode invitation, or a *diagnosis of breast cancer* in the previous screening process, are common features taken into account for a decision in favour of non-attendance of the woman to the episode.

Although the screening unit could elect to apply those rules independently in order to target those women likely not to attend the screening invitation, an algorithm could be more sensibly developed to perform this task. Furthermore, the variety of methods and derived rules are not the only difficulties encountered when attempting to predict attendance. Almost all of the generated models analysed incurred avoidable prediction errors, the most common being: i) positive prediction of attendance in cases where there is no invitation, or ii) a general prediction when there is no Townsend information available. These errors need to be targeted and solved before any serious attempt to predict the attendance of women to the screening invitation is formulated. The following section concentrates on the formulation of such an algorithm.

6.5 The artificial intelligence attendance (AI-ATT) algorithm

In the light of the previous observations, formulation of a predictive algorithm that will embrace the best of the models previously generated in a per-episode structure, given a batch of invited women, is needed.

6.5.1 The AI-ATT algorithm

This section concentrates on the formulation of a new algorithm based on artificial intelligence techniques for the prediction of the attendance of a particular woman to her screening invitation.

The AI-ATT algorithm is formulated as follows:

-Step 1 ***Loading a batch file.***

The batch (selection of women whose attendance it is of interest to predict) is extracted from the Breast Screening Unit database in a separate Microsoft compatible file.

-Step 2 ***Manipulation of data, decision on the Number of episodes (I) involved and generation of new variables.***

This step involves the manipulation of the extracted file in order to re-organise the data in terms of the screening history of each woman by

episode, and once achieved, to conclude the number of different screening episodes existing in the data under analysis.

For each one of the existing screening episodes, new variables are generated.

-Step 3 ***File partitioning by episode.***

For each woman in the re-organised and extended file, a decision has to be made as to which episode the last invitation is for.

This decision is based on the analysis of the values of the variable *Ageband* for each of the episodes in her screening history, as this variable will have values different from zero only when there has been an invitation for that episode. Particularly, the last episode at which she has received an invitation (last episode with *Ageband*≠*0*), will be determined. This will be her actual episode. The decision can be formulated on the basis of:

$$(Ageband_1 \neq 0) and \left(\left(\sum_{i=2}^{I} Ageband_i \right) = 0 \right)$$

for the first episode and

$$(Ageband_j \neq 0) and \left(\left(\sum_{i=j+1}^{I} Ageband_i \right) = 0 \right)$$

for all the other existing episodes.

The re-organised file will be further divided into sub-files, depending on the number of episodes that exist for a particular batch. These will then be

populated with the woman's history following her actual screening episode; i.e., a woman invited for the first time by the unit will be stored in the sub-file dedicated to the first episode, a woman who is being invited at her second screening episode in the sub-file for the second episode, and so on.

We shall label these files $A_1,\ldots, A_j,\ldots, A_I$, where *I* is the number of episodes involved.

-Step 4 ***File partitioning by Townsend deprivation score information.***

Once classification by screening episode has been performed, information relating to a woman's social deprivation must be determined, based on the area in which she was living at the point of invitation.

In order to achieve this, a mapping of the *Townsend deprivation score* with the *area code* is performed. The corresponding value of the score is stored in the variable *Townsref* (note that if such information is not available for a particular woman, the corresponding value of the variable will be empty).

With this information, further classification of the women is performed. This time, each one of the files A_j, corresponding to episode j, will be partitioned into sub-files for women with or without social deprivation information.
Such files are labelled as follows:

- A_j^{TS} for the sub-file of A_j with *Townsend social deprivation score* information, and,

- A_j^{PA} for the sub-file of A_j without *Townsend social deprivation score* information, which will therefore operate on information relating to *Postal area.*

Thus, the rule for this partitioning is expressed as:

For each episode file A_j, each woman in the file will be tested.

If she has a value stored in her corresponding variable for social deprivation information then the woman will be added to the file A_j^{TS}, otherwise, she will be added to the file A_i^{PA}.

This step finalises the process of allocating a woman with respect to the screening episode for which her attendance will be predicted, and the information available related to her residential area at invitation.

-Step 5 ***Predicting attendance for the episode.***

As previously discussed, it has been found that different ANN models perform differently for each screening episode in both data sets. Subsequently, each one of the sub-files obtained in the previous step needs to be submitted to different models in order to achieve an initial prediction for the woman to her

present invitation for the episode. It has also been noted that there is no significant difference in the results obtained by the generated models using the AI methods ANNP and RBFN.

Therefore, for each sub-file obtained, the *attendance* of each of the women involved will be predicted using the models generated by ANNP and RBFN for the corresponding type of sub-file.

Within each sub-file, the results of this prediction will be stored in four variables: two variables for each method, one storing the predicted value for attendance, and the second storing the confidence value for the prediction.

These variables may be referred to as:

^{ANNP}Att – the predicted attendance obtained with the ANNP generated model

^{RBFN}Att – the predicted attendance obtained with the RBFN generated model

$_{P}{}^{ANNP}$Att – the confidence value for the predicted attendance obtained with the ANNP generated model

$_{P}{}^{RBFN}$Att – the confidence value for the predicted attendance obtained with the RBFN generated model

-Step 6 ***Applying the classifier.***

In the previous step, two possible predictions have been obtained for the attendance of the woman to the present screening episode, but only one is necessary. In order to resolve this issue, a simple voting classifier is implemented.

This classifier compares the confidence value of the prediction obtained by both models, and votes in favour of the attendance prediction given by the model with higher confidence value.

Thus,

$A_{jn}{}^{k}Att$ – the predicted attendance for woman *n* in file $A_j{}^{k}$

Then the classifier can be expressed as:

If $\quad {}_{P}{}^{ANNP}Att > {}_{P}{}^{RBFN}Att \quad$ then $\quad A_{jn}{}^{k}Att = {}^{ANNP}Att$

Otherwise, $\quad A_{jn}{}^{k}Att = {}^{RBFN}Att$

This process is applied to each woman in each of the sub-files.

It should be mentioned that, although in this work this voting classifier has been implemented for use with only two models, it could be easily extended for use with several models.

-Step 7 ***Obtaining final results.***

Having obtained the predictive results for each woman in each sub-file, these are now re-assembled in the original file with the predictive values added. In other words, the sub-files will be concatenated again, but this time with only the general variables of the women, those related to her present episode and two new variables, the predicted attendance value and the confidence value of the predicted attendance.

The variables can thus be defined as:

Att – attendance of women to their corresponding screening episode

$_{P}$Att – the confidence value for the predicted attendance obtained

A_j^kAtt – the predicted attendance for all women in file A_j^k, where *k* can take values *TS* or *PA*

$_{P}A_j^k$Att – the confidence value for the predicted attendance for all women in file A_j^k, where *k* can take values *TS* or *PA*

Then the final variables Att and ${}_{P}Att$ can be expressed as,

$$Att = \bigcup_{i=1}^{I} \bigcup_{k} \left(A_i^{\,k} \, Att \right)$$

$${}_{P}Att = \bigcup_{i=1}^{I} \bigcup_{k} \left({}_{P}A_i^{k} \, Att \right)$$

Fig. 6.5 shows the implementation stream of the AI-ATT algorithm using visual programming under a Clementine environment.

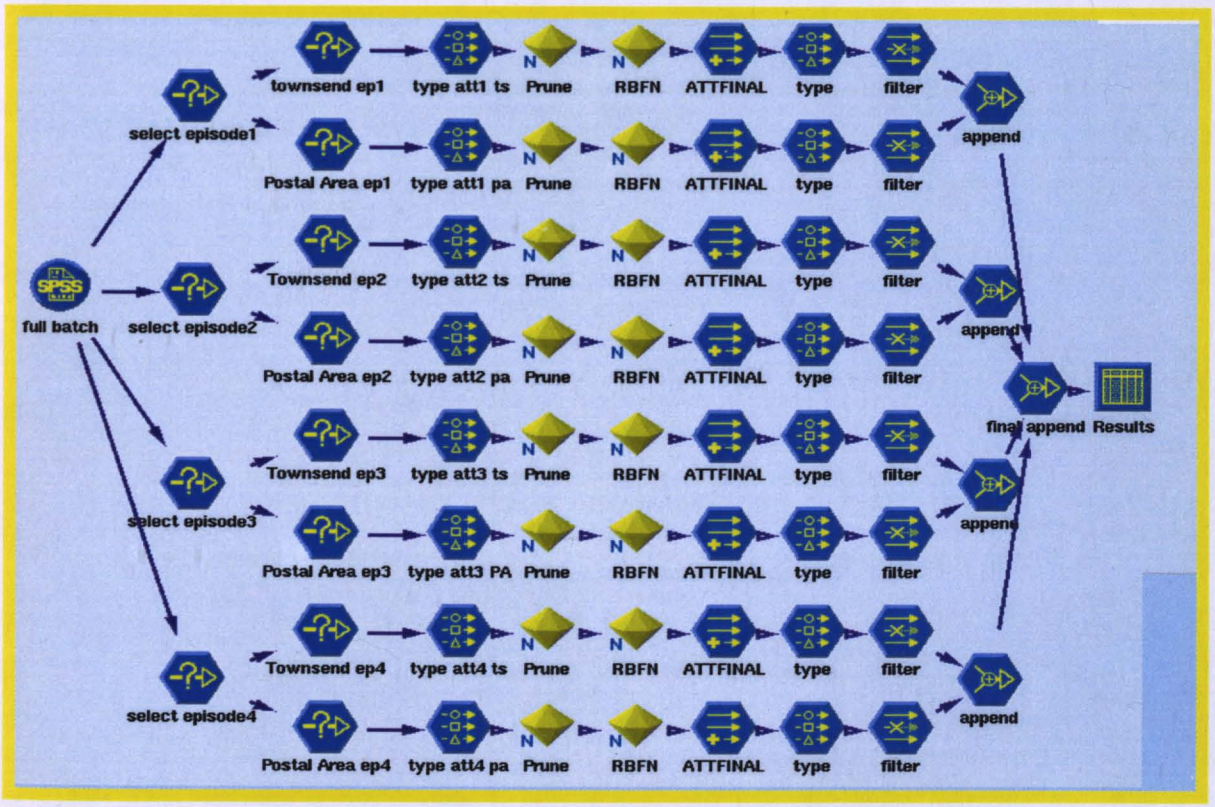


Fig. 6.5 Visual programming in Clementine of the AI-ATT algorithm

The AI-ATT algorithm has been designed to deal with errors intrinsic to women attending different episode numbers in the same batch, and those with no social

deprivation information. Results using this approach are compared with those achieved using the models independently.

It is important to note that, when running the models without separation of the data according to episode, the LR, ANNP and RBFN predict not only the episode for which the woman has been invited lately (her current episode), but also all of the previous episodes too. These results are extremely confusing and point to the need for data separation before the application of the models. This process is included as part of the AI-ATT algorithm.

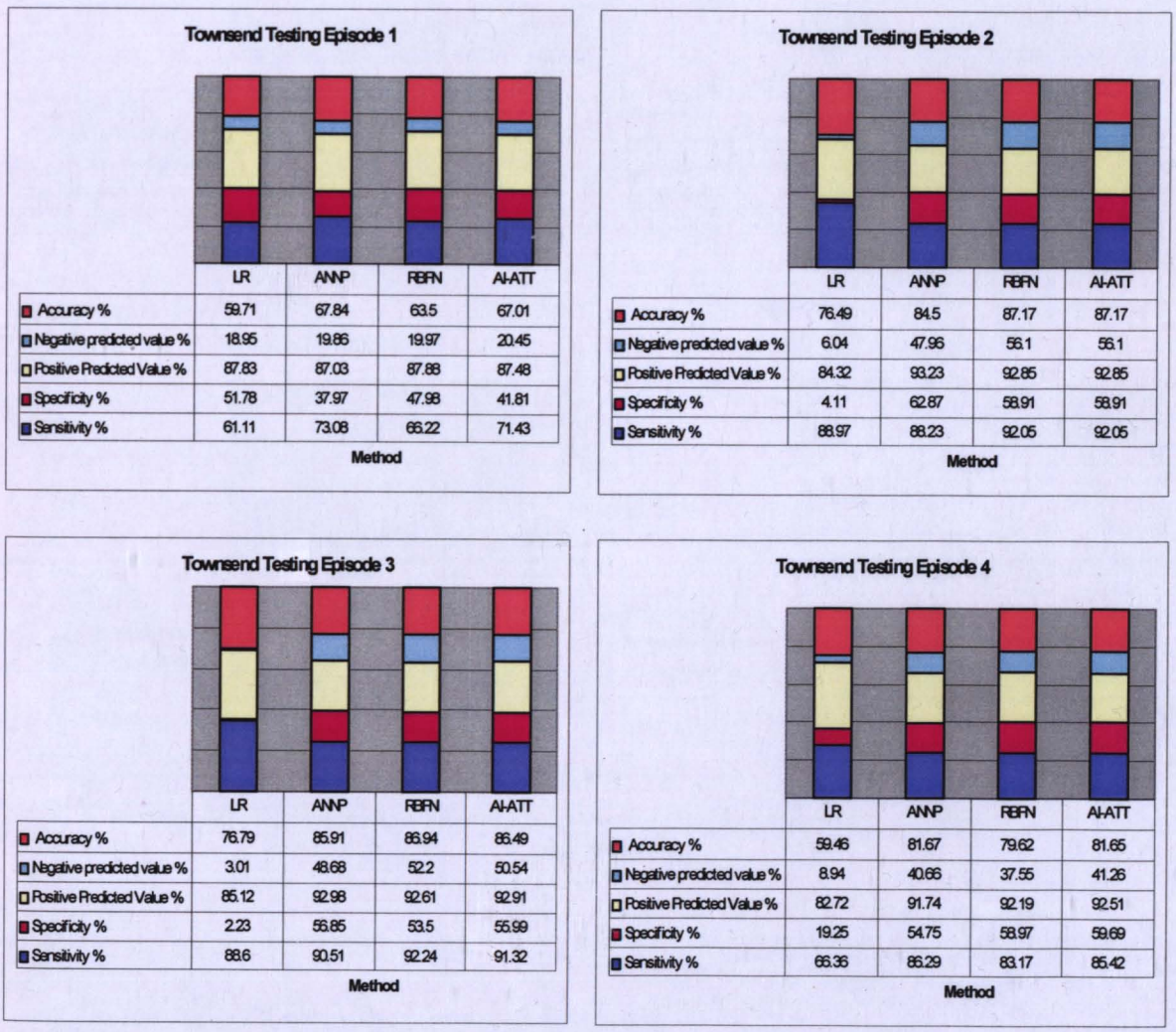


Fig. 6.6. Performance comparison for Townsend test data set (%) All methods.

Bearing this in mind, the results discussed below are those obtained by each of the methods assuming that the data has been separated beforehand.

Figures 6.6 and 6.7 depict the results obtained for all the methods for all the episodes with the test data set (data not previously seen by the models).

Fig. 6.7. Performance comparison for Postal area test data set (%) All methods

Table 6.IX shows results obtained using the individual models, compared with those using the AI-ATT algorithm for the test data set.

Table 6.IX AI-ATT PERFORMANCE COMPARISON VS. BEST OTHER METHOD

Performance Parameter	Episode	Test data set	Best performing method (LR, ANNP or RBFN)		AI-ATT performance value (%)	Difference between methods (AI-ATT - Best other) (%)
			Method	value (%)		
Accuracy	1	Postal area	ANNP	63.05	63.24	0.19
		Townsend	ANNP	67.84	67.01	-0.83
	2	Postal area	ANNP	87.84	87.76	-0.08
		Townsend	RBFN	87.17	87.17	0.00
	3	Postal area	RBFN	85.41	84.74	-0.67
		Townsend	RBFN	86.94	86.49	-0.45
	4	Postal area	RBFN	69.60	66.44	-3.16
		Townsend	ANNP	81.67	81.65	-0.02
Negative predicted value	1	Postal area	ANNP	15.61	15.70	0.09
		Townsend	RBFN	19.97	20.45	0.48
	2	Postal area	ANNP	55.46	55.09	-0.37
		Townsend	RBFN	56.10	56.10	0.00
	3	Postal area	RBFN	44.68	42.96	-1.72
		Townsend	RBFN	52.20	50.54	-1.66
	4	Postal area	RBFN	35.48	33.95	-1.53
		Townsend	ANNP	40.66	41.26	0.60
Positive predicted value	1	Postal area	RBFN	87.09	86.84	-0.25
		Townsend	RBFN	87.88	87.48	-0.40
	2	Postal area	ANNP	92.70	92.72	0.02
		Townsend	ANNP	93.23	92.85	-0.38
	3	Postal area	ANNP	92.76	92.87	0.11
		Townsend	ANNP	92.98	92.91	-0.07
	4	Postal area	ANNP	88.59	87.78	-0.81
		Townsend	RBFN	92.19	92.51	0.32
Specificity	1	Postal area	LR	51.10	37.19	-13.91
		Townsend	LR	51.78	41.81	-9.97
	2	Postal area	RBFN	53.54	53.52	-0.02
		Townsend	ANNP	62.87	58.91	-3.96
	3	Postal area	ANNP	52.99	53.96	0.97
		Townsend	ANNP	56.85	55.99	-0.86
	4	Postal area	ANNP	74.10	64.59	-9.51
		Townsend	RBFN	58.97	59.69	0.72
Sensitivity	1	Postal area	ANNP	67.26	67.48	0.22
		Townsend	ANNP	73.08	71.43	-1.65
	2	Postal area	ANNP	93.27	93.14	-0.13
		Townsend	RBFN	92.05	92.05	0.00
	3	Postal area	RBFN	90.32	89.33	-0.99
		Townsend	RBFN	92.24	91.32	-0.92
	4	Postal area	RBFN	73.17	66.93	-6.24
		Townsend	ANNP	86.29	85.42	-0.87

Analysing the results for each performance parameter in the table, it can be observed that, in terms of accuracy, the AI-ATT algorithm achieves results within less than 0.84% difference of those achieved by the best of the other methods (0.63% average), except for the *Postal area data set* in the fourth episode where the difference is 3.16 % in favour of the RBFN algorithm.

In terms of negative predictive value, the AI-ATT performs within 1.72%, or less, difference of the results achieved by the best of the other methods (0.51% average). Moreover, in five out of the eight different cases, the AI-ATT algorithm performs within 0.6% difference or less.

Analysed from the point of view of the positive predictive value, the AI-ATT algorithm performs within 0.81%, or less, of the results obtained by the best performing model, with an average difference of 0.18%.

The specificity parameter shows very inconsistent results with differences ranging from a 13.91%, in favour of the model generated by the Logistic Regression method for the *Postal area data set* in the first episode, up to 0.97% in favour of the AI-ATT for the *Postal area data set* in the third episode. The average of the difference for the eight possible cases is 4.57%. However this average is achieved by a very polarised result, four cases averaging a 9.34% difference, opposite to the other four cases which average a difference of 0.2%.

The average difference of the AI-ATT algorithm with respect to the best of the other methods is 1.32%. But six out of the eight cases have less than 0.92% difference (with an average of 0.45%).

Thus, although not always achieving the best values, the AI-ATT has a consistently good performance for all the possible cases with respect to all parameters. Moreover, the performance achieved by the algorithm is not overthrown by any of the other models as can be seen in the graph depicted in Fig. 6.8, where the only case when the AI-ATT achieves higher average difference than the other methods is for the specificity parameter and, even then, the difference between all methods is smaller than 0.62%. Note that the results for the LR method have been excluded from Fig. 6.8. This is due to the relatively high differences obtained for this method.

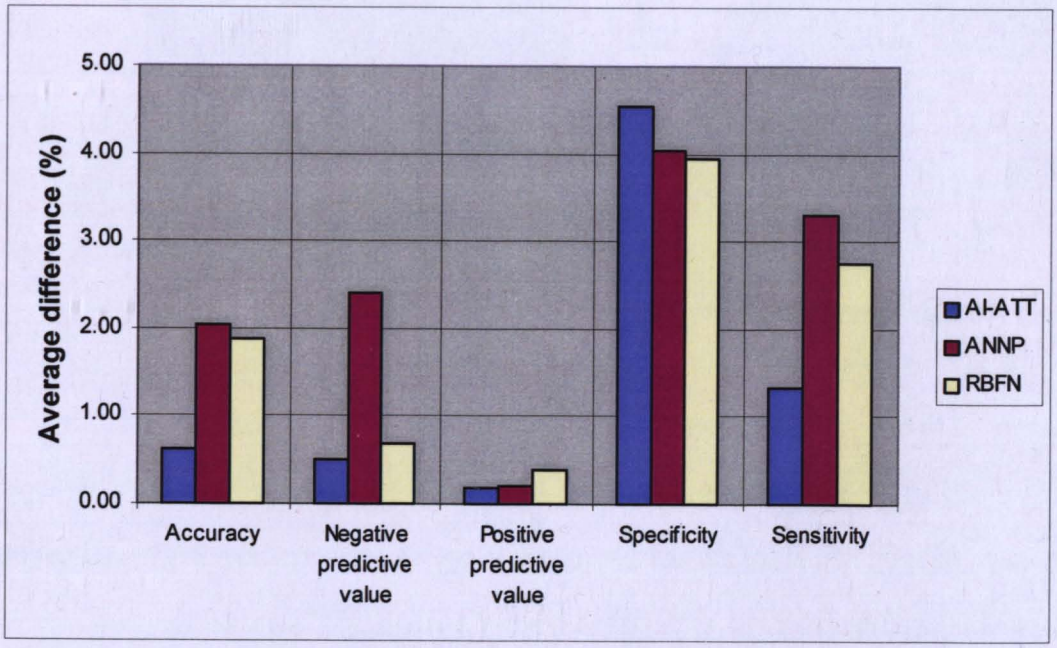


Fig. 6.8 Comparison by method of the average differences with respect to the best performing model

These results are extremely favourable towards the use of the AI-ATT algorithm in the prediction of attendance of women to breast screening.

Additionally, analysis of the validation set results (Fig. 6.9), also show the AI-ATT to have better performance in general than any of the other methods when all the parameters are taken into account. It particularly achieves a higher negative predictive value (non-attendance prediction) for all the episodes in both data sets (*Townsend* and *Postal Area*); the highest (90.23%) is obtained in the second episode for the *Postal Area data set*. The LR method has the poorest performance for all episodes (only in the first one, when there is no previous screening history of the women, is it comparable with the other methods).

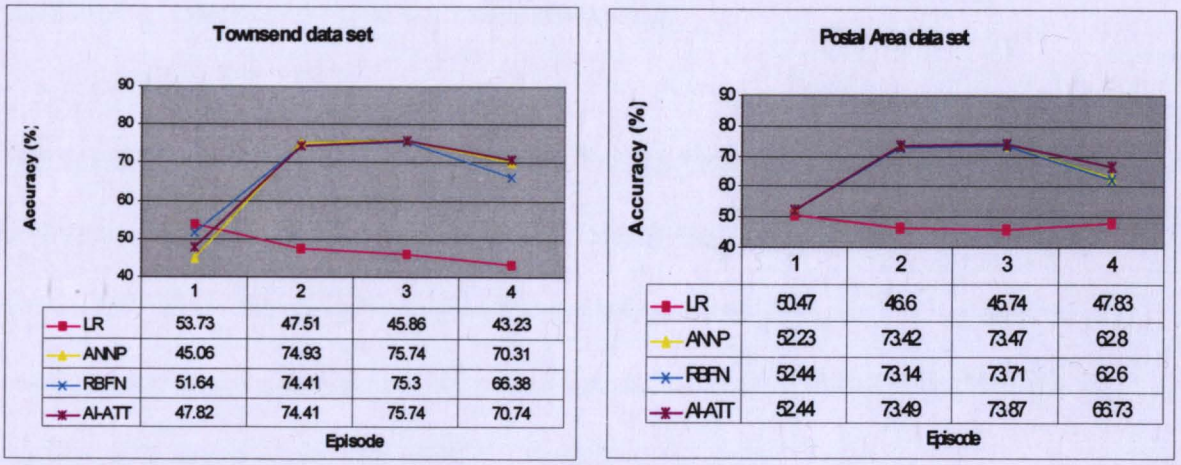
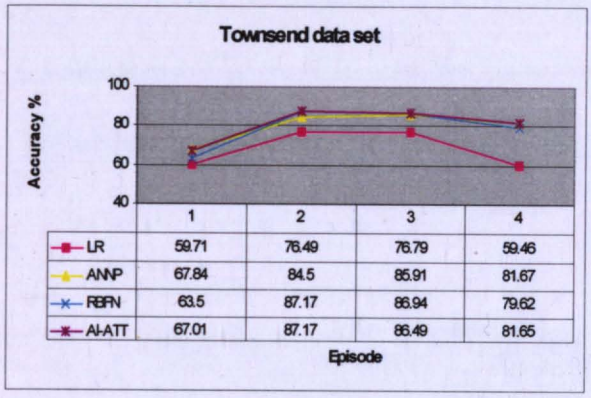


Fig. 6.9. Accuracy results for validation data set.

The results obtained for the validation set have been corroborated again by a close inspection of the performance of the accuracy of the methods for the test data (Fig. 6.10), where the highest accuracy of the AI-ATT (87.76%) is obtained in the second episode for the *Postal Area data set*.



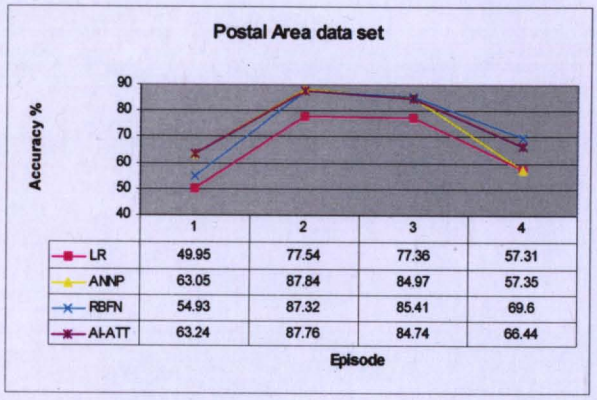


Fig 6.10. Accuracy results for test data set

In general, for the test data, the best results are obtained for the *Townsend data set*, thus advocating the use of the *Townsend deprivation score* information in the prediction of *attendance* at the screening programme.

The classification in Step 6 of the AI-ATT algorithm achieved at least the same performance as those for the ANNs for each one of the screening episodes. However, the AI-ATT algorithm may be seen as a more robust alternative, as it achieves overall good results for all episodes regardless of the differences in individual data sets for each episode as previously discussed.

When predicting *attendance*, not only is high accuracy important, but also detection of those women who are likely *not to attend*, so that appropriate measures may be put in place to improve *attendance*. In this case, the specificity is an important performance parameter in the prediction of *attendance*, i.e., the ability of the method to predict negative values. The impressive positive predictive values of the methods have been previously discussed. Here, a high value indicates that those women identified as attendees have a high probability of being "real" attendees. The specificity allows us to determine how good the model is at predicting women who will *not* attend. It has been

already pointed out that this was the weakest parameter in the AI-ATT performance. Nevertheless, Fig. 6.11 shows the behaviour of the predictive methods discussed with respect to their specificity for validation and test data sets.

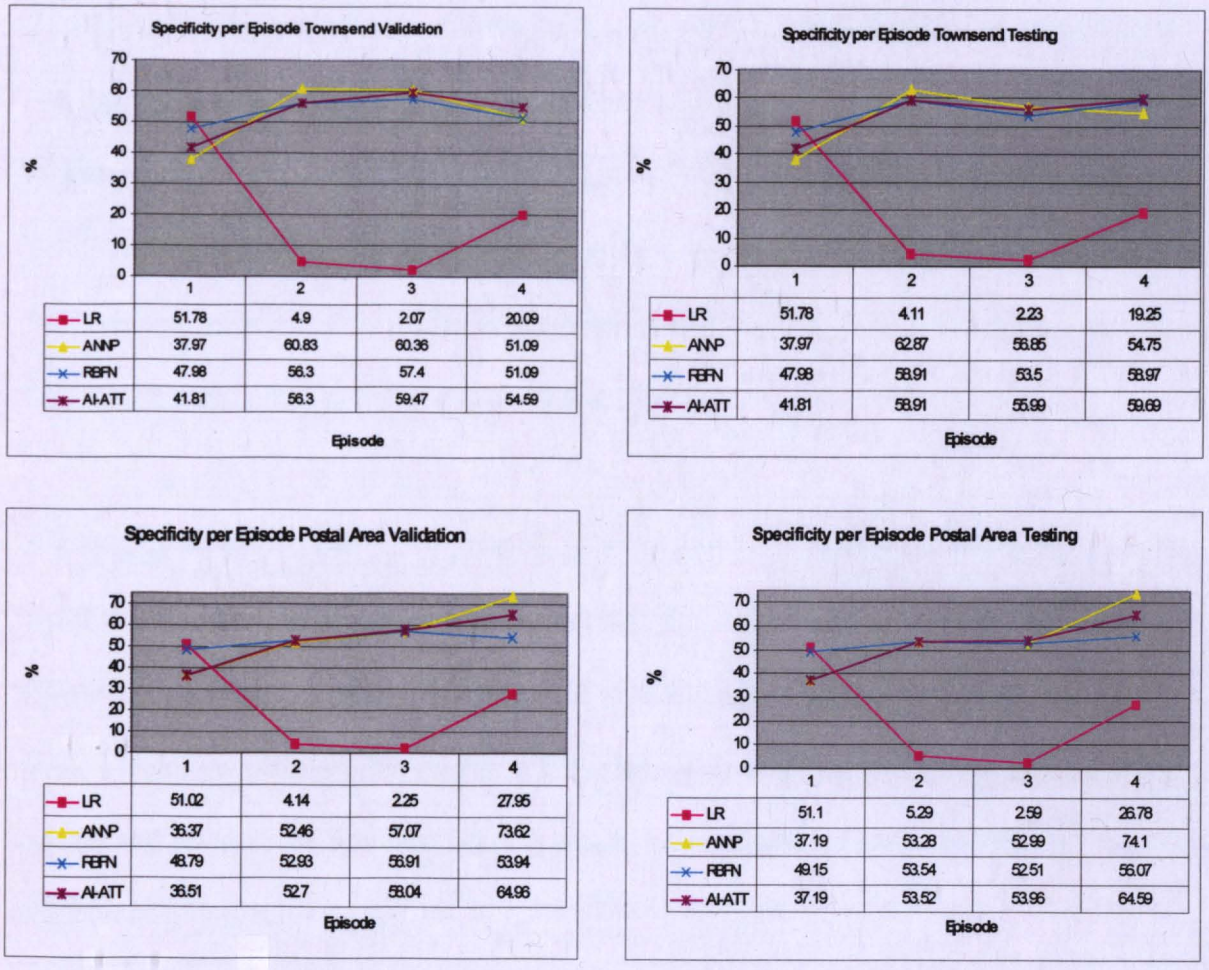


Fig 6.11 Specificity per episode for the attendance models

By closely studying the specificity, it can be shown that the LR method also fails to adequately detect such women (non-attenders). However, we should not overlook the fact that this latter method achieved the best detection of non-attenders for the first episode. On the other hand, the AI models, including the AI-ATT algorithm, detected at least half of such women for episodes 2, 3 and 4, and an average of 40% of them for the first episode. The graphs also show a tendency of the algorithms to increase their ability

to detect non-attenders as the number of the episode increased, i.e., the models showed a high potential to improve as more information of the screening history of the women became available.

Looking at the impact of those results for the Breast Screening Programme, the method proposed gives the programme a predictive tool able not only to predict efficiently those women who are likely to attend, but also, the capability to detect a reasonable number of those who are likely not to attend. This gives the programme the opportunity of targeting such women efficiently in order to increase attendance in the first instance and to decrease mortality due to breast cancer in the long term.

Drawing up an exhaustive list of possible interventions with respect to non-attenders that could point to an improvement in attendance falls outside the scope of this study. However, the literature points to the use of second appointment letters, personal phone calls, invitations through the women's GPs, localisation of screening vans nearer their homes and awareness, amongst others, which have proved to make an impact on the attendance of women.

Chapter 7

Prediction of Screening Variation

In the case of women responding to the screening invitation, it is beneficial to determine the approximate date on which they will actually attend the programme. This simply translates into a prediction of *screening variation.*

7.1 Topology of the artificial neural network models

The performance of a particular AI algorithm is known to depend closely on the parameters used in training the models. A particular algorithm can achieve either adequate or disastrous results depending on such parameters. Hence it is important to properly select and track changes in those parameters when refining the search towards the achievement of an appropriate model.

Although far from exhaustive, several architectures have been experimentally used in this analysis, with a wide and careful selection of parameters. The results in Tables 7.I and 7.II are those generated by the best models.

Table 7.I NUMBER OF NEURONS FOR NON-BALANCED BY SV CLASSES MODELS

Episode	Data Set	Method							
		ANNP				RBFN			
		Input	Layer 1	Layer 2	Output	Input	Layer 1	Output	η
1	Townsend	4	20	15	17	4	20	17	0.475
	Postal Area	3	4	1	17	3	20	17	0.433
2	Townsend	12	20	15	17	12	20	17	0.645
	Postal Area	11	20	15	17	11	20	17	0.608
3	Townsend	14	17	12	17	14	20	17	0.648
	Postal Area	13	16	8	17	13	20	17	0.556
4	Townsend	14	3	2	17	*			
	Postal Area	13	1	1	17	*			

*RBFN has not been able to be trained for the 4th episode

Table 7.II NUMBER OF NEURONS FOR BALANCED BY SV CLASSES MODELS

Episode	Data Set	Method							
		ANNP				RBFN			
		Input	Layer 1	Layer 2	Output	Input	Layer 1	Output	η
1	Townsend	4	17	12	17	4	20	17	0.438
	Postal Area	3	20	15	17	*			
2	Townsend	12	2	2	17	12	20	17	0.563
	Postal Area	11	7	6	17	11	20	17	0.486
3	Townsend	14	15	14	17	14	20	17	0.592
	Postal Area	13	18	11	17	13	15	17	0.559
4	Townsend	14	20	15	17	*			
	Postal Area	13	20	15	17	13	20	17	0.479

* The method has not been able to be trained

The parameters used for training the AI methods are summarised in Tables 7.III and 7.IV:

Table 7.III GENERAL PARAMETERS FOR ANNP

Parameters	Value
Hidden rate	0.15
Hidden persistence	6
Input rate	0.15
Input persistence	4
Persistence	100
Overall Persistence	3
α	0.9
Initial η	0.3
η decay	30
High η	0.1
Low η	0.01

Table 7.IV GENERAL PARAMETERS FOR RBFN

Parameters	Value
Number of RBF clusters	20
Persistence	30
α	0.9
η	computed automatically
RBF overlap	1.0

The results in Tables 7.I and 7.II allow the conclusion that, for the ANNP models, although the number of neurons in each layer varies with the episode and data sets, the two-hidden layers structure is the one achieving better predictive models. Meanwhile, for the RBFN method, although the number of neurons in the hidden layer does not vary, the value of the learning rate (η) varies with respect to the episode and data sets.

7.2 Comparison of results

This section concentrates on the analysis of the results obtained for the three methods of analysis, namely Logistic Regression, ANN Pruning and RBFN when predicting *screening variation*. An appropriate method should have not only high accuracy, but also a high capacity for correctly predicting positive values (sensitivity).

The results obtained are summarised in Tables E.I – E.XVI (Appendix E), and Fig.7.1 gives a graphical schema for those models trained with balanced data with respect to the screening variation classes.

A close analysis of the results shown in Fig. 7.2, highlights the lack of results for two cases for the RBFN algorithm (particularly for episode 1, *Postal area data set* and episode 4, *Townsend data set*). This is due to the impossibility of obtaining a generated model from the training process.

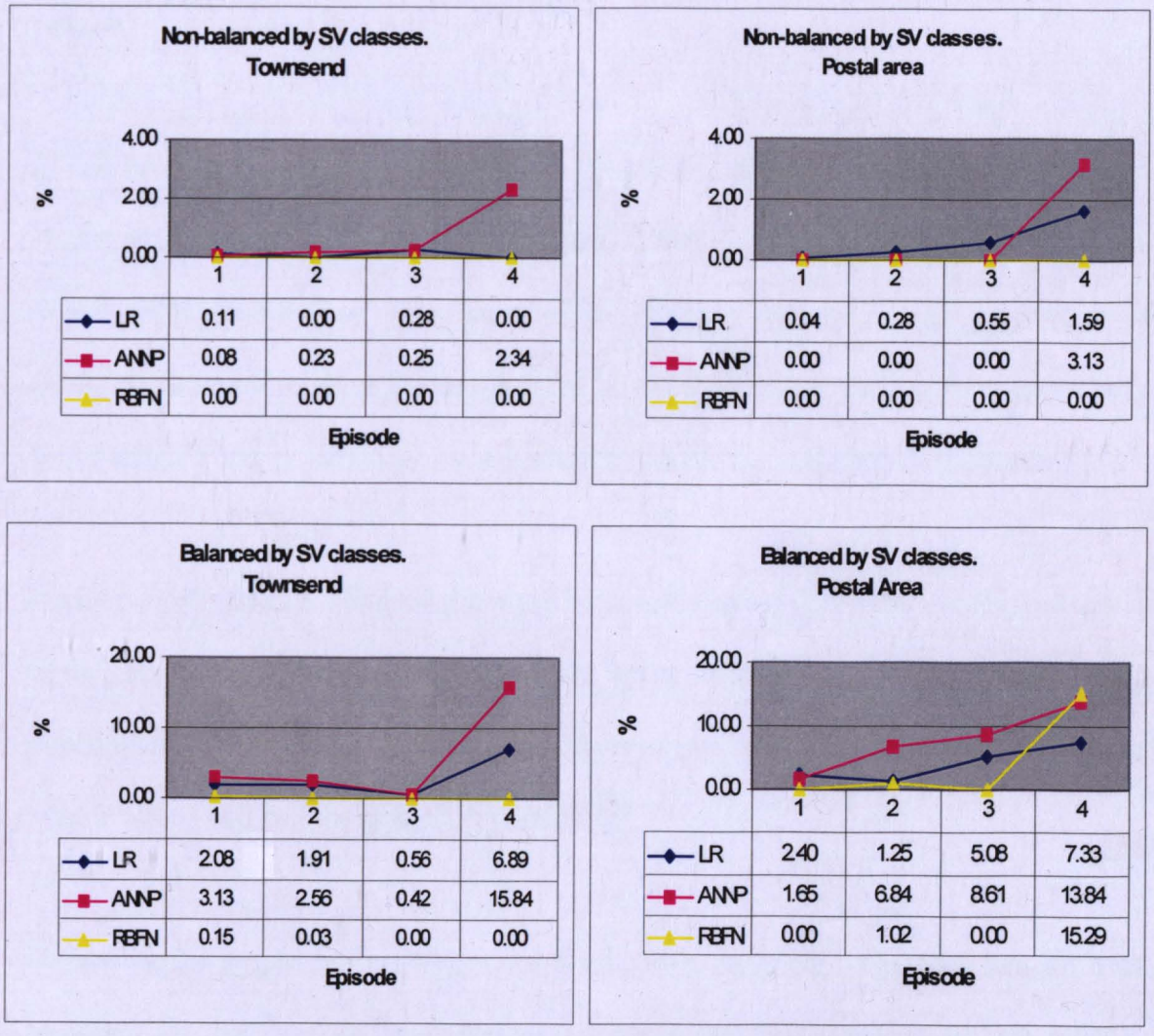


Fig. 7.1 Sensitivity results comparison for the test data set.

It can also be seen that those models achieving better accuracy perform worse in terms of sensitivity (correct prediction of screening variation), as demonstrated by the test data set in the third episode with *Postal area* information. Although the RBFN achieves an

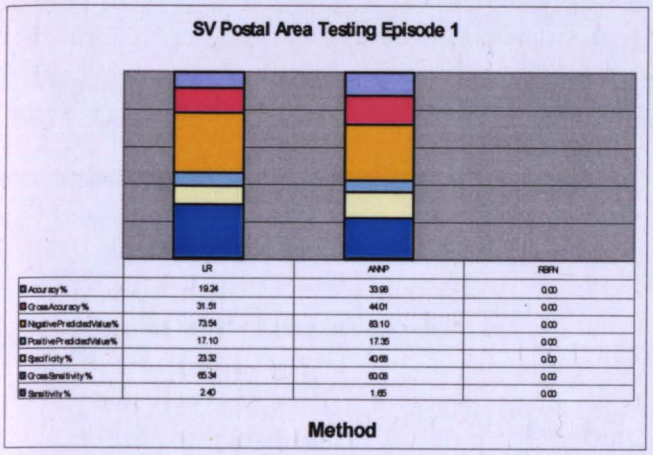
accuracy of 76.79%, its sensitivity was nil. On the other hand, the ANNP model, only achieve 47.81% accuracy, but its sensitivity was 8.61%.

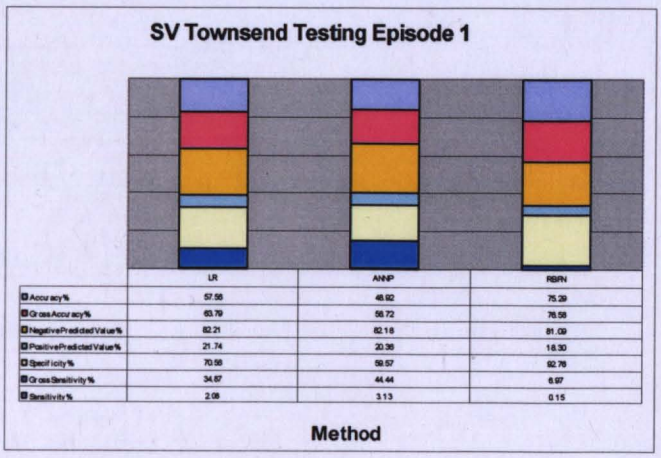
The graphs depicted in Fig. 7.1 show the results of the sensitivity achieved by the three methods analysed for predicting *screening variation* for all the generated predictive models.

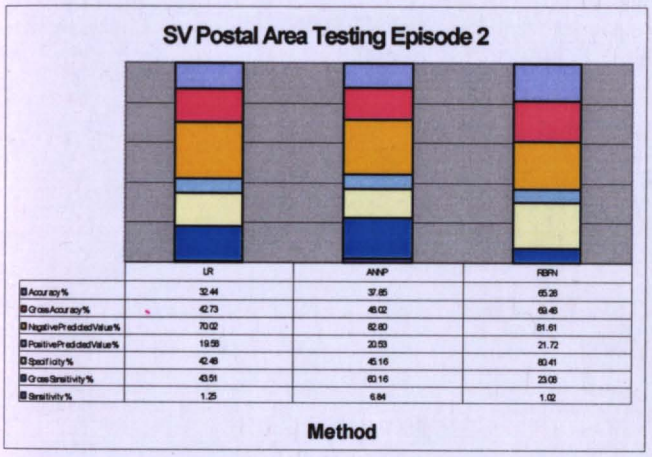
As can be seen from the graphs, although the models trained with *balanced data* with respect at screening variation classification, achieve better performance than those trained with *non-balanced data*, none of the methods achieve a high percentage of correct detection of positive values. In other words, the methods fail to correctly classify the existing *screening variation*, even when they could have detected its existence.

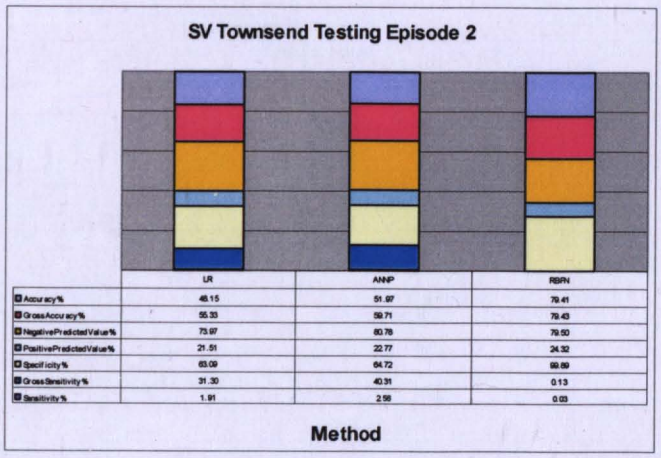
In summary, the results obtained show the failure of the three methods to detect a correct classification of *screening variation*, in the cases when it occurs. The possibility of detecting occurrence or non-occurrence of *screening variation* (independently of its classification) will now be studied in detail.

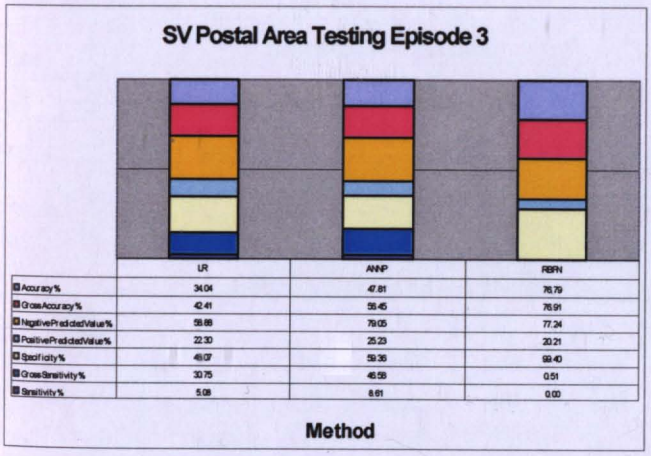
Comparing the graphs in Fig. 7.3 (where blanks cells mean that the model has not been generated for that case), with respect to the *non-balanced* and *balanced* data models, although the *non-balanced* model is observed to achieve gross accuracy in the band of 80% for the majority of the cases, it fails to detect *screening variation*. Its higher gross sensitivity is 12.18% and 14.58% for the *Townsend* and *Postal Area data sets*, respectively.

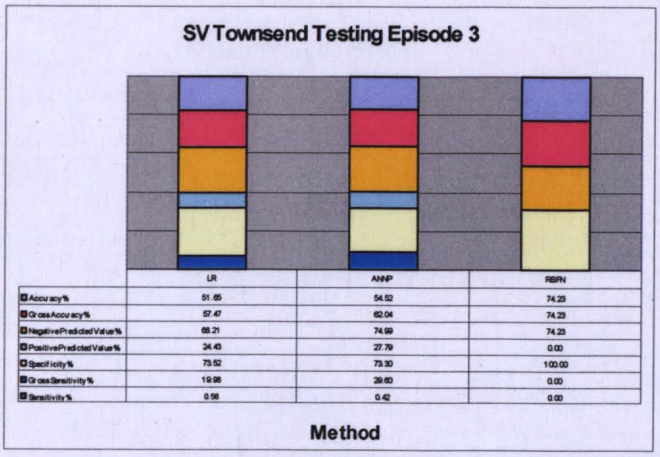


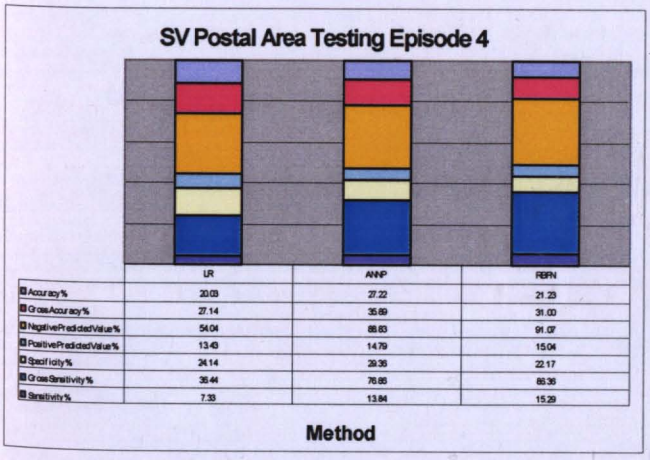












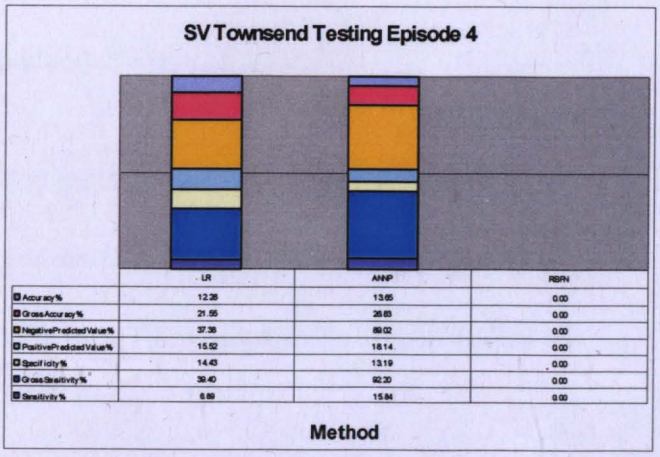


Fig.7.2 Comparison by methods for screening variation prediction. (Balanced by SV classes)

A close analysis of the results shown in Fig. 7.2, highlights the lack of results for two cases for the RBFN algorithm (particularly for episode 1, *Postal area data set* and episode 4, *Townsend data set*). This is due to the impossibility of obtaining a generated model from the training process.

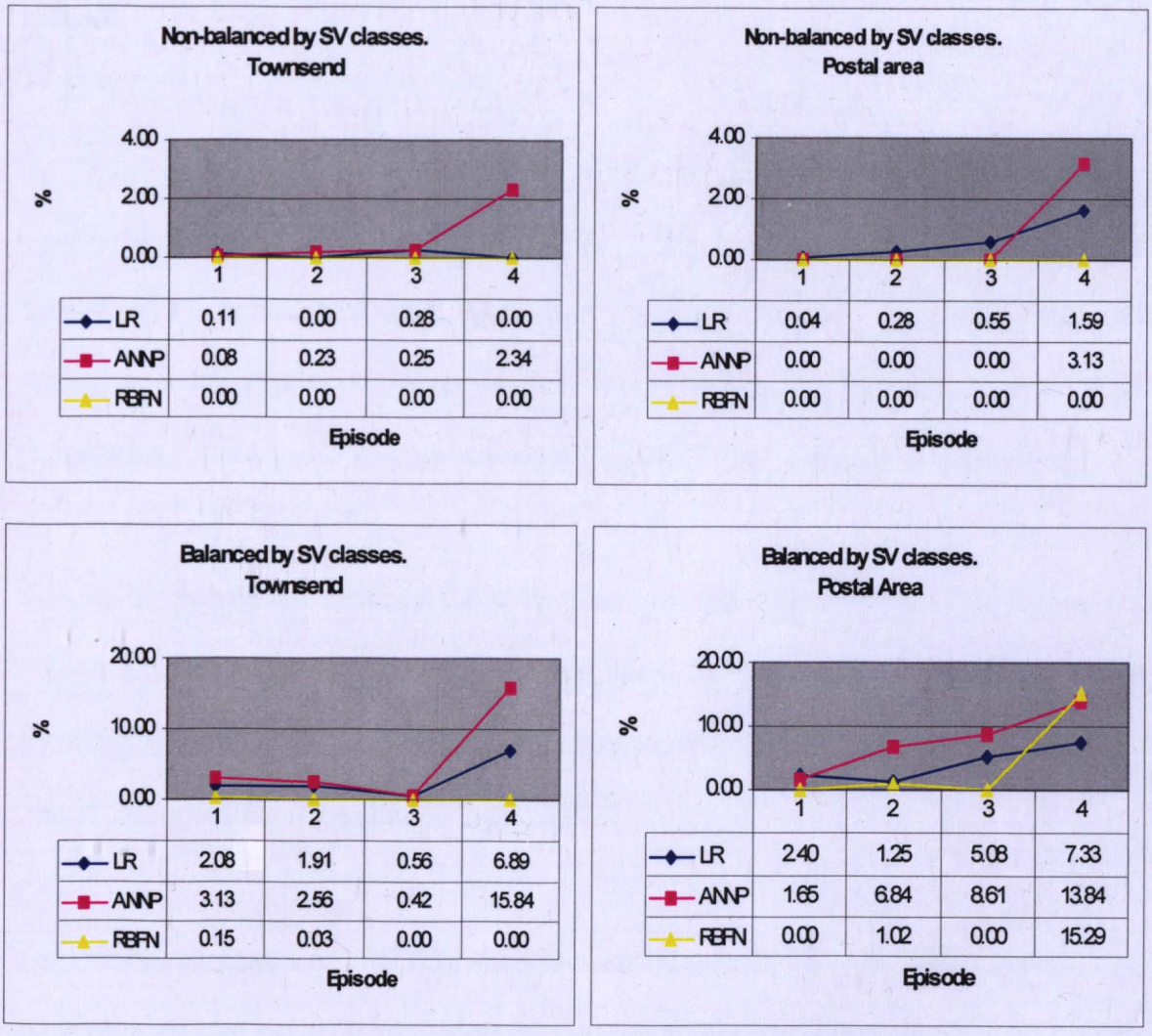


Fig. 7.1 Sensitivity results comparison for the test data set.

It can also be seen that those models achieving better accuracy perform worse in terms of sensitivity (correct prediction of screening variation), as demonstrated by the test data set in the third episode with *Postal area* information. Although the RBFN achieves an

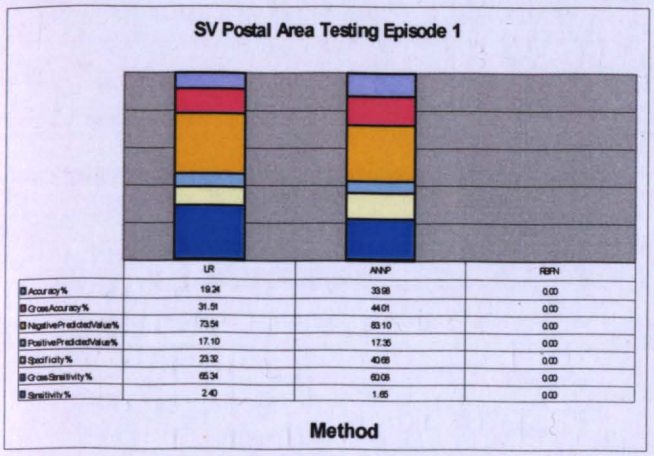
accuracy of 76.79%, its sensitivity was nil. On the other hand, the ANNP model, only achieve 47.81% accuracy, but its sensitivity was 8.61%.

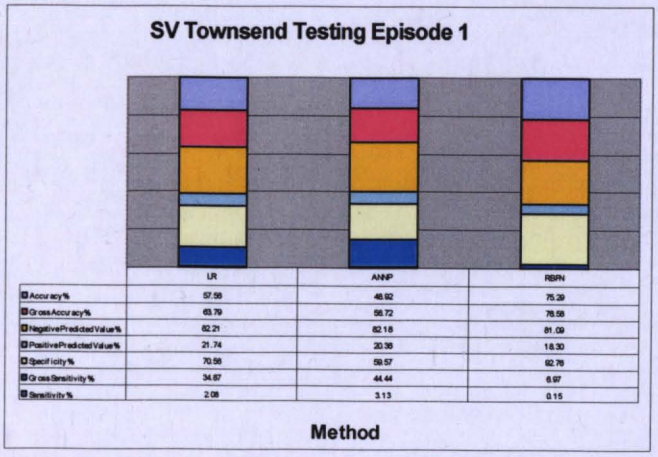
The graphs depicted in Fig. 7.1 show the results of the sensitivity achieved by the three methods analysed for predicting *screening variation* for all the generated predictive models.

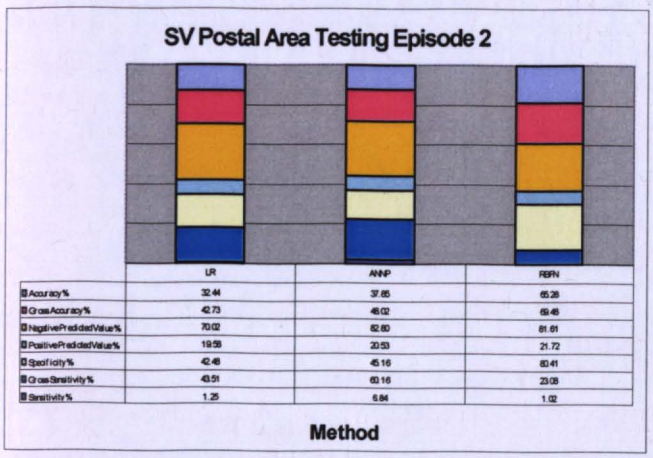
As can be seen from the graphs, although the models trained with *balanced data* with respect at screening variation classification, achieve better performance than those trained with *non-balanced data*, none of the methods achieve a high percentage of correct detection of positive values. In other words, the methods fail to correctly classify the existing *screening variation*, even when they could have detected its existence.

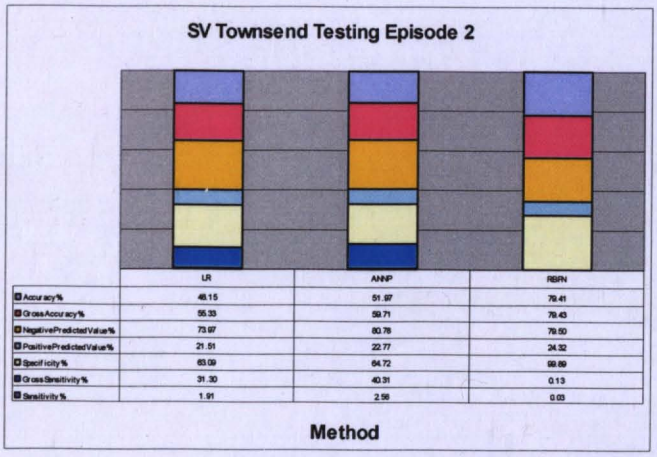
In summary, the results obtained show the failure of the three methods to detect a correct classification of *screening variation*, in the cases when it occurs. The possibility of detecting occurrence or non-occurrence of *screening variation* (independently of its classification) will now be studied in detail.

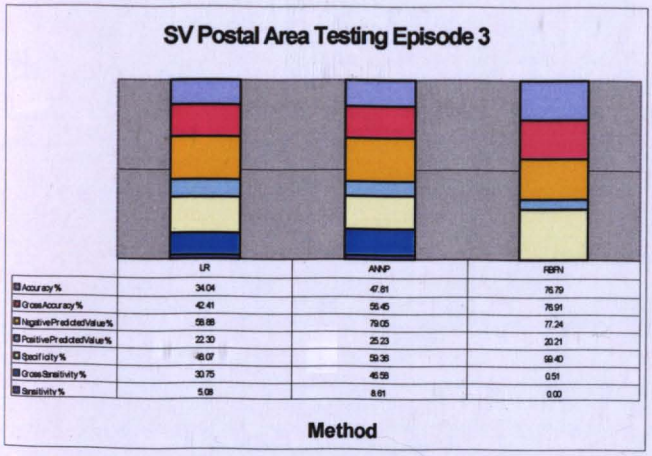
Comparing the graphs in Fig. 7.3 (where blanks cells mean that the model has not been generated for that case), with respect to the *non-balanced* and *balanced* data models, although the *non-balanced* model is observed to achieve gross accuracy in the band of 80% for the majority of the cases, it fails to detect *screening variation*. Its higher gross sensitivity is 12.18% and 14.58% for the *Townsend* and *Postal Area data sets*, respectively.

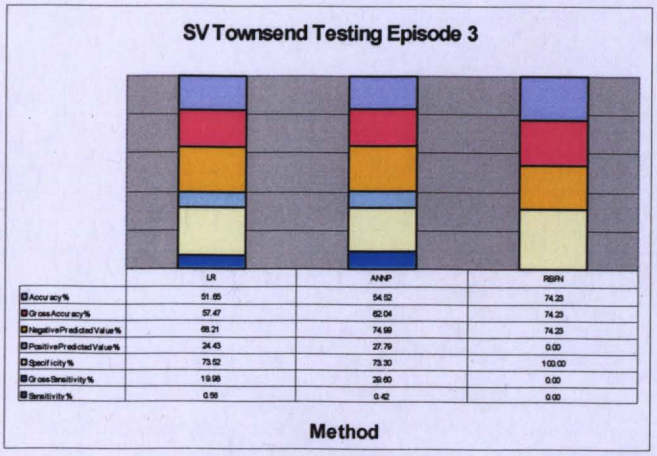


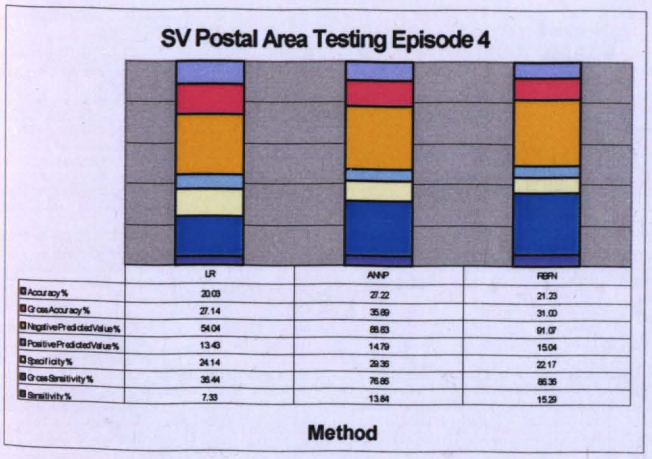












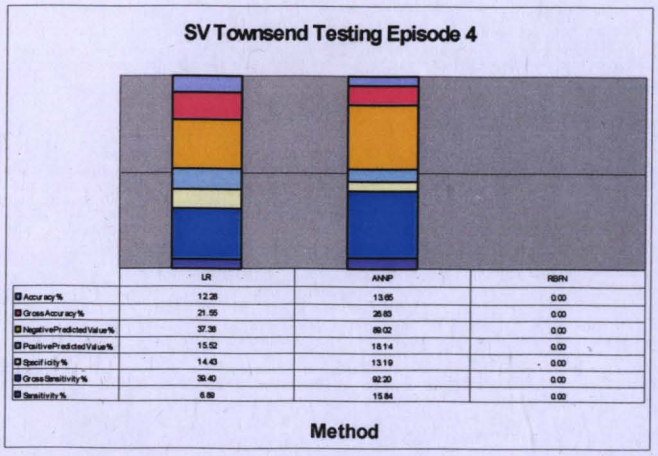
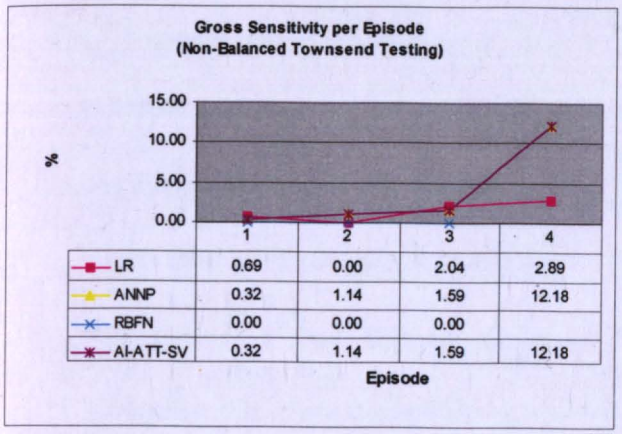
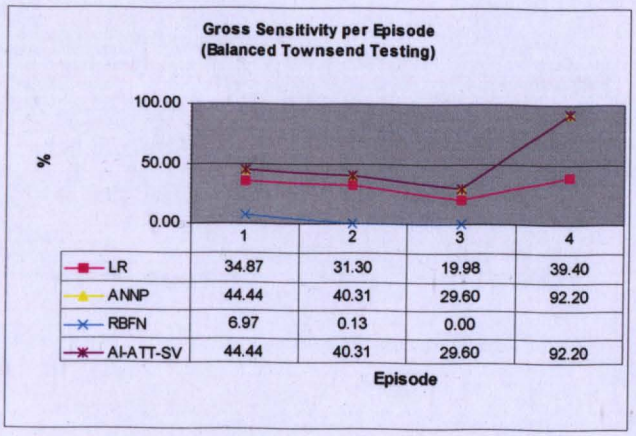
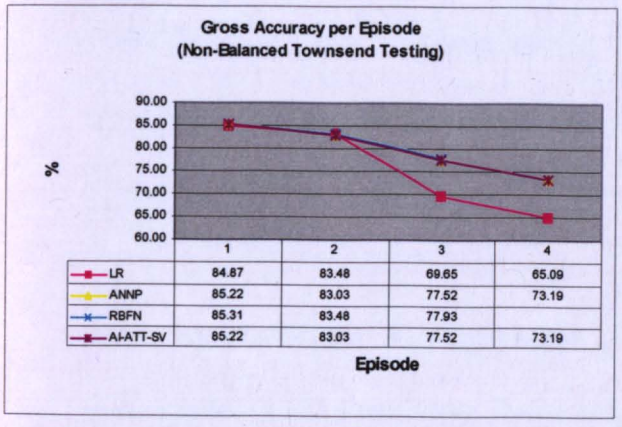
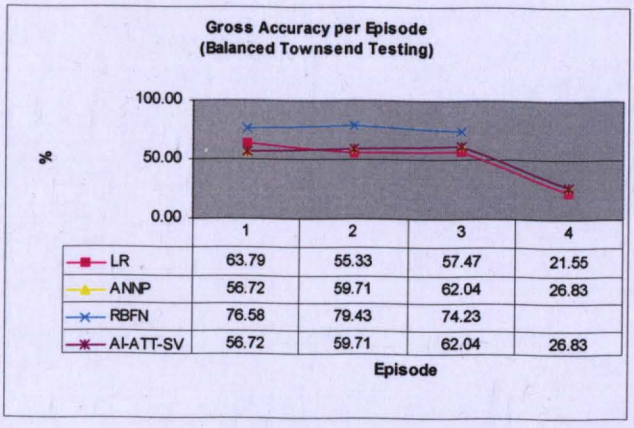


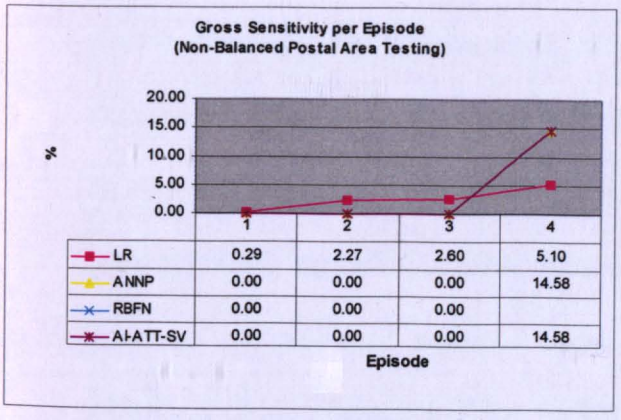
Fig.7.2 Comparison by methods for screening variation prediction. (Balanced by SV classes)

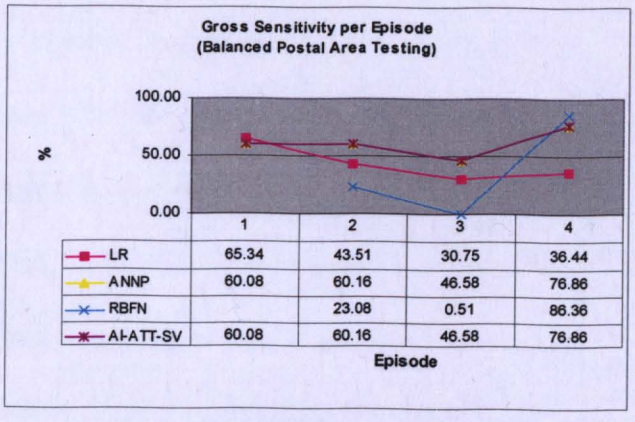


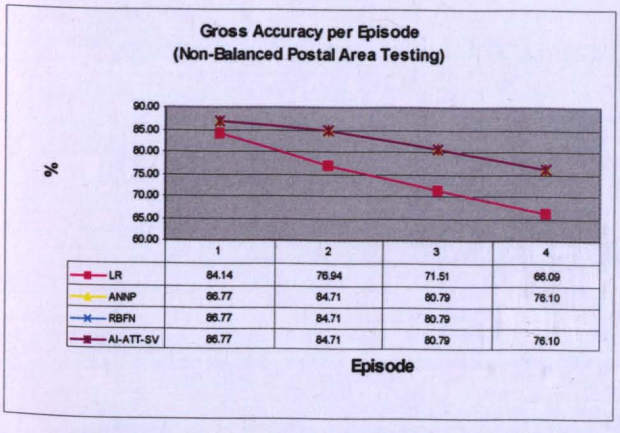












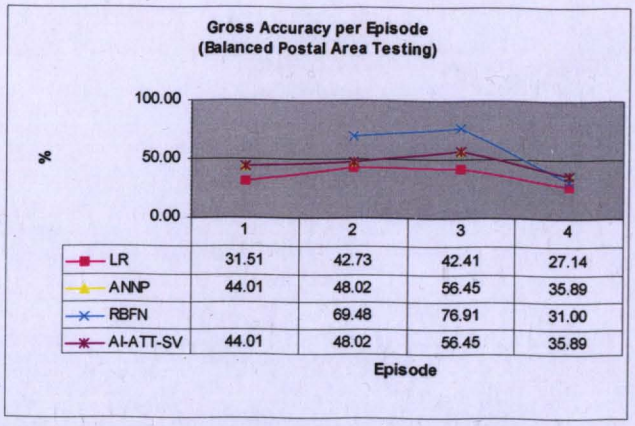


Fig 7.3. Gross sensitivity and accuracy comparison by screening variation predictive method

In contrast, balanced models achieve a low percentage of gross accuracy (minimum of 21.55% obtained with the LR model for the fourth episode, and maximum of 79.43% obtained with the RBFN for the second episode), but has higher gross sensitivity than the non-balanced data models (maximum of 92.2% obtained with the ANNP for the fourth episode), i.e., they achieve a better performance in detecting women undergoing screening variation. Nevertheless, the gross sensitivity values obtained by these models vary with the episode, and whether the analysis is in relation to the *Townsend* or *Postal Area* sets. The best performances in general can clearly be concluded to be obtained from the ANNP models.

Thus, those methods achieving a better performance in gross accuracy terms are the ones with worse gross sensitivity. These results lead to the conclusion that, although the models developed fail to correctly predict the non-occurrence of *screening variation* or its accurate length, they are still able to predict occurrence of *screening variation* in general terms. Presently, such models can be used as a tool for detecting women likely to incur *screening variation* in order to take further steps towards their *attendance* on time, or as an indication of possible empty appointment slots when planning screening sessions. Further discussion of the possible causes of, and solutions to, this failure of the predictive models is presented in section 7.3.

7.3 Extension of the AI-ATT algorithm

Once the *attendance* of a woman at screening has been predicted, the prediction of the *screening variation* can be carried out. Given the robustness of the AI-ATT algorithm for predicting *attendance*, a logical extension is to use a similar technique for the

prediction of *screening variation*. Therefore, a new algorithm is formulated, as a natural extension of the AI-ATT, which not only predicts *attendance* but also predicts *screening variation*.

7.3.1 The AI-ATT-SV algorithm

The extended algorithm is formulated as follows:

-Steps 1 to 6 are identical to those of the AI-ATT algorithm.

-Step 7 *Screening Variation sub-file generation.*

As discussed in previous sections, the predictors of *screening variation* and the architecture of predictive models, vary with the episode number and the availability of social deprivation data. Thus, in order to be able to predict *screening variation* an appropriate partitioning of the sub-files is needed.

In this step the screening variation sub-files which will be labelled SV_j^k will be generated.

Here, j=1,...,*I*, *I* is the number of episodes, and k takes values *TS* or *PA*.

The women identified as being likely to attend constitute the records in these sub-files. Note that those women predicted not to attend are not included in the SV_j^k sub-files, and consequently, no screening variation prediction is

generated for them. The identification of a woman predicted to attend is achieved as follows:

From Step 6, there are $2I$ sub-files A_j^k such as j=1,…,I and k takes values *TS* or *PA*.

A_{js}^k Att represent the predicted attendance of woman *s* in sub-file A_j^k

Then if the woman is predicted as being likely to attend the screening invitation, i.e., A_{js}^k Att=1, then she will be transferred to sub-file SV_j^k. Otherwise, she remains in sub-file A_j^k.

At the end of this process, there will be $4I$ sub-folders:

$2I$ sub-folders A_j^k containing those women predicted not to attend, and
$2I$ sub-folders SV_j^k containing women predicted to attend the present screening invitation.

-Step 8 ***Predicting Screening Variation for the episode.***

As discussed in previous sections, different predictive models perform differently for each screening episode and residential area information available. Therefore, each of the sub-files SV_j^k needs to be submitted to a different model in order to predict the *screening variation* of the women. Based on the results discussed previously, models used for the prediction will

be those generated by the AI methods, ANNP and RBFN, trained with data balanced by screening variation classes.

Within each sub-file, the results of the *screening variation* prediction will be stored in four variables; two of them for each method, one storing the predicted value, and the second storing the confidence value for the prediction.

Let these variables be defined as,

^{ANNP} ScrV – the predicted screening variation obtained with the ANNP generated model

^{RBFN} ScrV – the predicted screening variation obtained with the RBFN generated model

$_{P}{}^{ANNP}$ScrV – the confidence value for the predicted screening variation obtained with the ANNP generated model

$_{P}{}^{RBFN}$ScrV – the confidence value for the predicted screening variation obtained with the RBFN generated model

-Step 9 ***Applying the classifier for screening variation.***

In the previous step, two possible predictions were obtained for the *screening variation* of the woman to the present screening episode, but only one is

necessary. As before (section 6.5.1), a simple voting classifier has been implemented.

As with the *attendance* classifier, this classifier compares, for each independent woman, the confidence value of the prediction obtained by both models, and votes in favour of the *screening variation* prediction given by the model with higher confidence value.

Let, $SV_{js}{}^{k}ScrV$ represent the predicted screening variation for woman *s* in file SV_{j}^{k}.

Then the classifier can be expressed as:

If $_{P}{}^{ANNP}ScrV > {}_{P}{}^{RBFN}ScrV$ then $SV_{js}{}^{k}ScrV = {}^{ANNP}ScrV$

Otherwise, $SV_{js}{}^{k}ScrV = {}^{RBFN}ScrV$

This process is applied to each woman in each of the sub-files.

-Step 10 ***Obtaining final results.***

Once the predictive results for each woman in each sub-file are obtained, these should now be re-assembled in the original file with the predictive values added.

Based on the following variable definitions (assuming that k assumes a value *TS* or *PA* where applicable):

Att – attendance of the women to their corresponding episode.

ScrV – screening variation of the women in their corresponding episode.

$_{P}$Att –confidence value for the predicted attendance obtained.

$_{P}$ScrV – confidence value for the predicted screening variation obtained.

A_j^kAtt – predicted attendance for all women in file A_j^k.

SV_j^kAtt – predicted attendance for all women in file SV_j^k.

A_j^kScrV – predicted screening variation for all women in file A_j^k.

Note that this variable will be a null value as there is no screening variation prediction for those women with non-attendance prediction.

SV_j^kScrV – predicted screening variation for all women in file SV_j^k.

$_{P}A_j^k$Att – confidence value for the predicted attendance for all

women in file A_j^k.

$_PSV_j^kAtt$ – confidence value for the predicted attendance for all women in file SV_j^k.

$_PA_j^kScrV$ – confidence value for the predicted screening variation for all women in file A_j^k.

Note that this variable will be a null value as there is no screening variation prediction for those women with non-attendance prediction.

$_PSV_j^kScrV$ – confidence value for the predicted screening variation for all women in file SV_j^k.

then the final variables *Att*, $_P$*Att*, *ScrV* and $_P$*ScrV* can be expressed as,

$$Att = \left(\bigcup_{i=1}^{I} \bigcup_{k} \left(A_i^k \, Att \right) \right) \cup \left(\bigcup_{i=1}^{I} \bigcup_{k} \left(SV_i^k \, Att \right) \right)$$

$$_P\,Att = \left(\bigcup_{i=1}^{I} \bigcup_{k} \left({}_P A_i^k \, Att \right) \right) \cup \left(\bigcup_{i=1}^{I} \bigcup_{k} \left({}_P SV_i^k \, Att \right) \right)$$

$$ScrV = \left(\bigcup_{i=1}^{I} \bigcup_{k} \left(A_i^k \, ScrV \right) \right) \cup \left(\bigcup_{i=1}^{I} \bigcup_{k} \left(SV_i^k \, ScrV \right) \right)$$

$$_{P}ScrV = \left(\bigcup_{i=1}^{I} \bigcup_{k} \left({}_{P}A_i^k ScrV \right) \right) \bigcup \left(\bigcup_{i=1}^{I} \bigcup_{k} \left({}_{P}SV_i^k ScrV \right) \right)$$

As a first approach to the prediction of *screening variation*, the model generated by the ANN Pruning, trained with balanced by screening variation classes data, is the one used in the proposed algorithm given the stability of this model with respect to performance for all the episodes and data sets. Thus, the classifier explained in step 9 has not been tested. This, however, should not be taken as a final decision. When more input information becomes available and a model is developed able to successfully predict *screening variation*, then further analysis of the performance of the methods should be carried out, and perhaps a voting method involving different models should be implemented in this step of the algorithm. Such work has not been performed in this project due to the lack of appropriate data to carry out an accurate prediction of *screening variation*. The Breast Screening Programme is strongly advised to start collecting data relative to the socio-economic factors of the women. Based on studies carried out worldwide, particularly in the USA, data relating to educational background, income, ethnicity and religion, amongst others, have proved to be closely correlated to *attendance* of women to screening. Hence, such factors are assumed to be quite likely to be strongly related to occurrence of *screening variation*. Furthermore, based on studies carried out in the UK, it is proposed that additional information is collected e.g. the distance from a woman's home to the screening facility (fixed or mobile unit). Other parameters likely to be of interest are information on car ownership, the use of public transport to attend an appointment, childcare provisions and employment status. A field study should be set up, as a first step, in order to prove whether these parameters assist in

the prediction of *screening variation* in the Breast Screening Programme in the UK, and to evaluate the extra cost that collecting such parameters would involve.

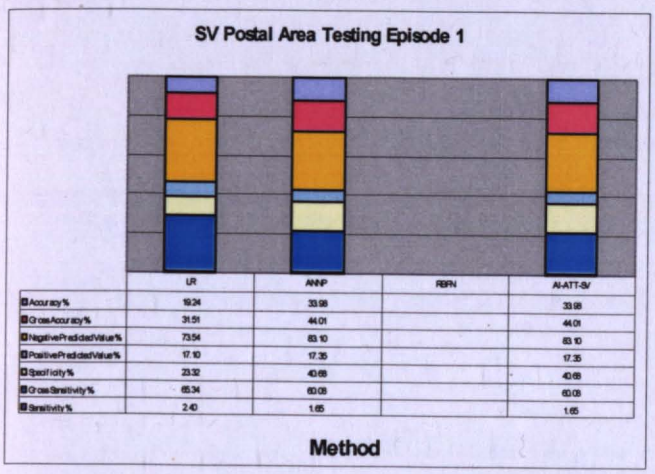
The following conclusions can be drawn by focusing our attention now on the results obtained in this project and taking into account that the results of the algorithm will coincide with the individual results of each ANNP model for each episode. The results obtained with the AI-ATT-SV algorithm with respect to those of the methods analysed in section 6.1 are compared in Fig 7.4, but only taking into account the balanced by *screening variation* classes models in the test data set.

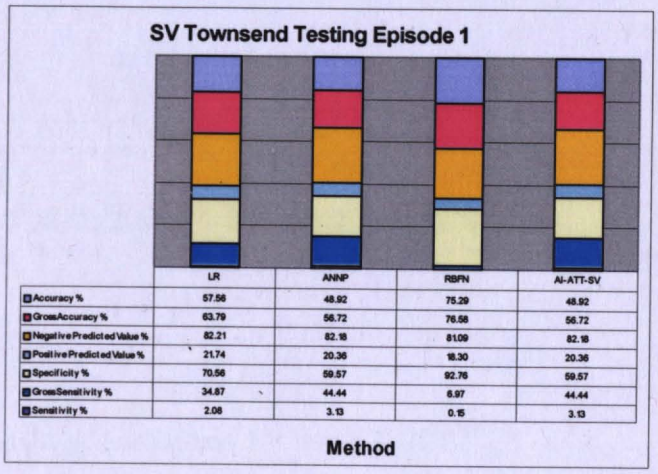
First, the results obtained for the first episode for the *Townsend data set* differed notably from those obtained for the *Postal Area data set*. The *Townsend data set* resulted in higher values for all the parameters analysed than the *Postal Area*, except for the gross sensitivity (capacity of the system of detecting occurrence of screening variation). For the gross sensitivity, the highest *Townsend data set* value achieved by all models (including AI-ATT-SV) is 44.44%, while for the *Postal Area set* the highest value achieved is 65.34% (obtained from the LR model). The AI-ATT-SV achieves 60.08%.

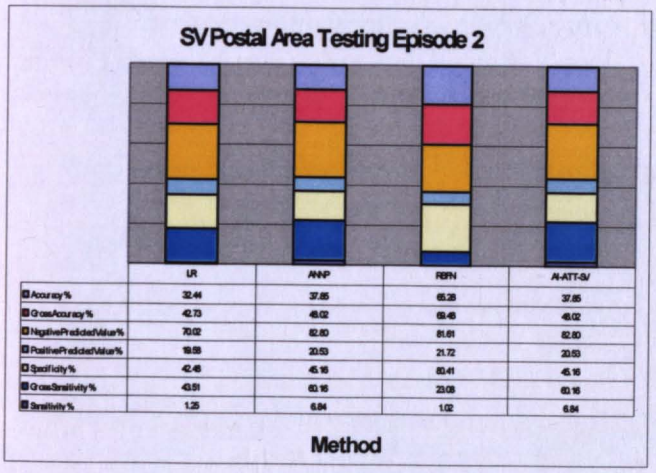
Such a difference was not observed for any of the other screening episodes. This observation points to the importance of the social deprivation information of the woman in the prediction of *screening variation* at the first screening episode, where no history is yet available of the woman's behaviour towards screening.

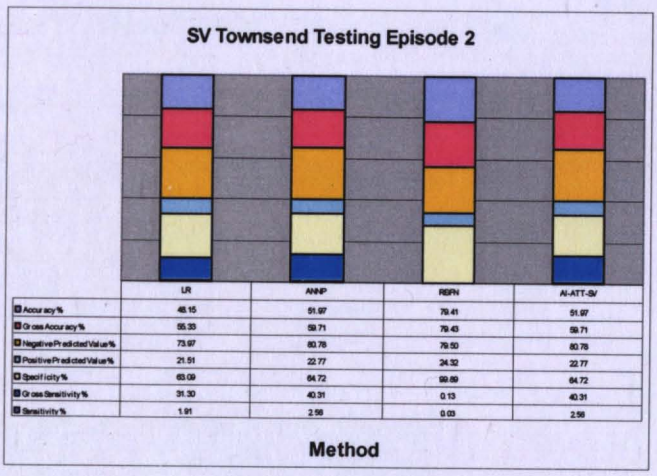
For the *Townsend data set*, the AI-ATT-SV achieved its highest accuracy and gross accuracy for the third episode (54.52% and 62.04%, respectively). On the other hand, the

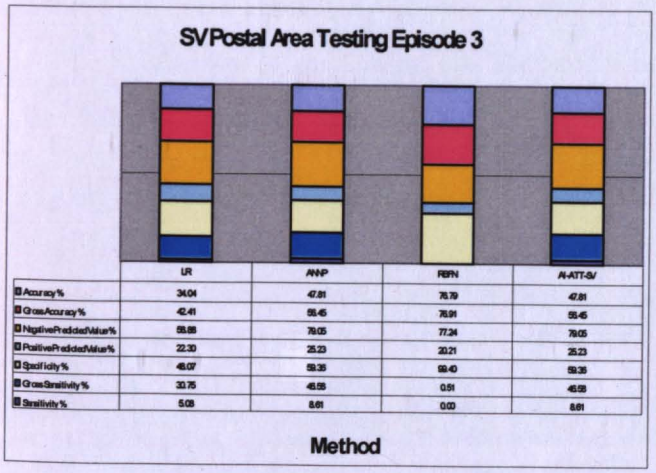
specificity (capacity of the system in predicting no screening variation) achieved its highest value in the third episode (73.3%). Therefore, the best accuracy was obtained when the best specificity was obtained. This result is logical given that the number of women incurring *screening variation* was very small compared with those who attended on the day of invitation (as a rule 80% or more of the women who attend do so on the invitation date). Nevertheless, the objective is to detect those 20% who do not, so the parameters of greater importance become the sensitivity and gross sensitivity. The fourth episode showed the highest sensitivity and gross sensitivity (15.84% and 92.20%, respectively). The conclusion is that, even when the algorithm does not predict highly accurately the correct classification of a woman's *screening variation* or a high accuracy in general, it is still able to predict whether a woman will incur a *screening variation.*

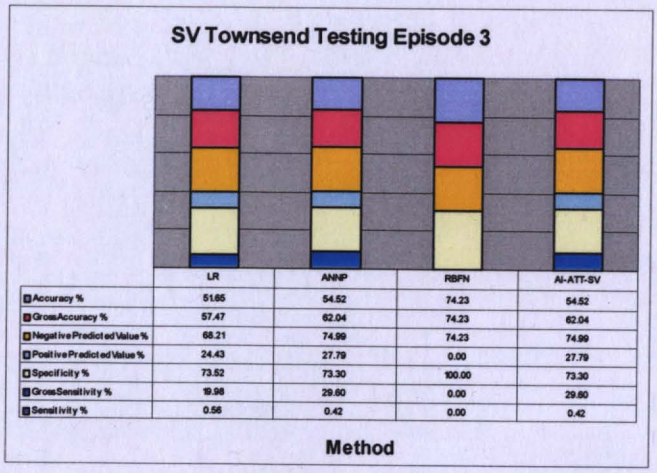


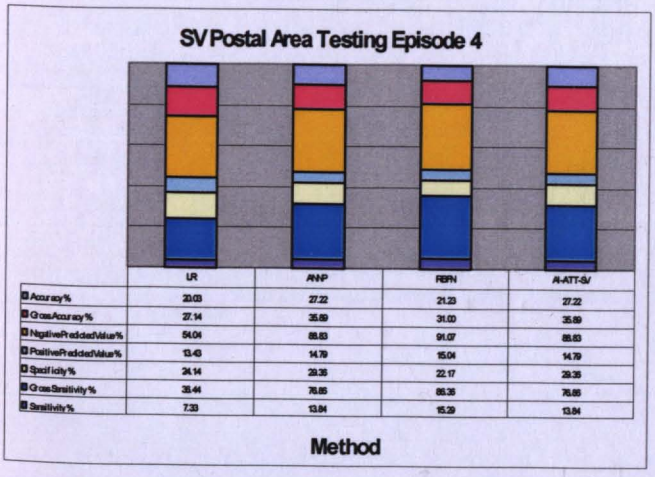












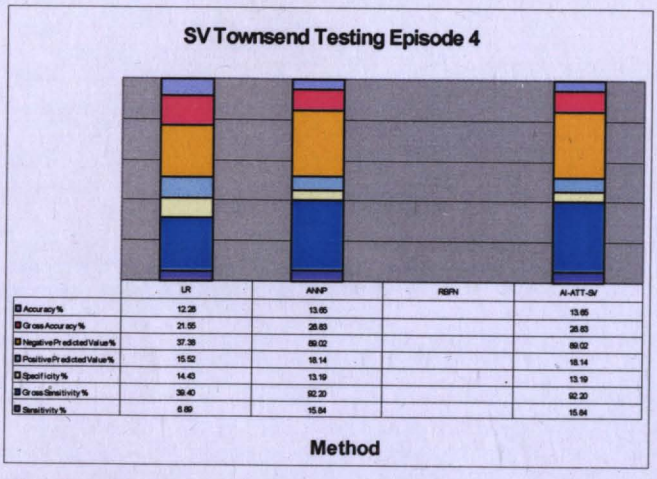
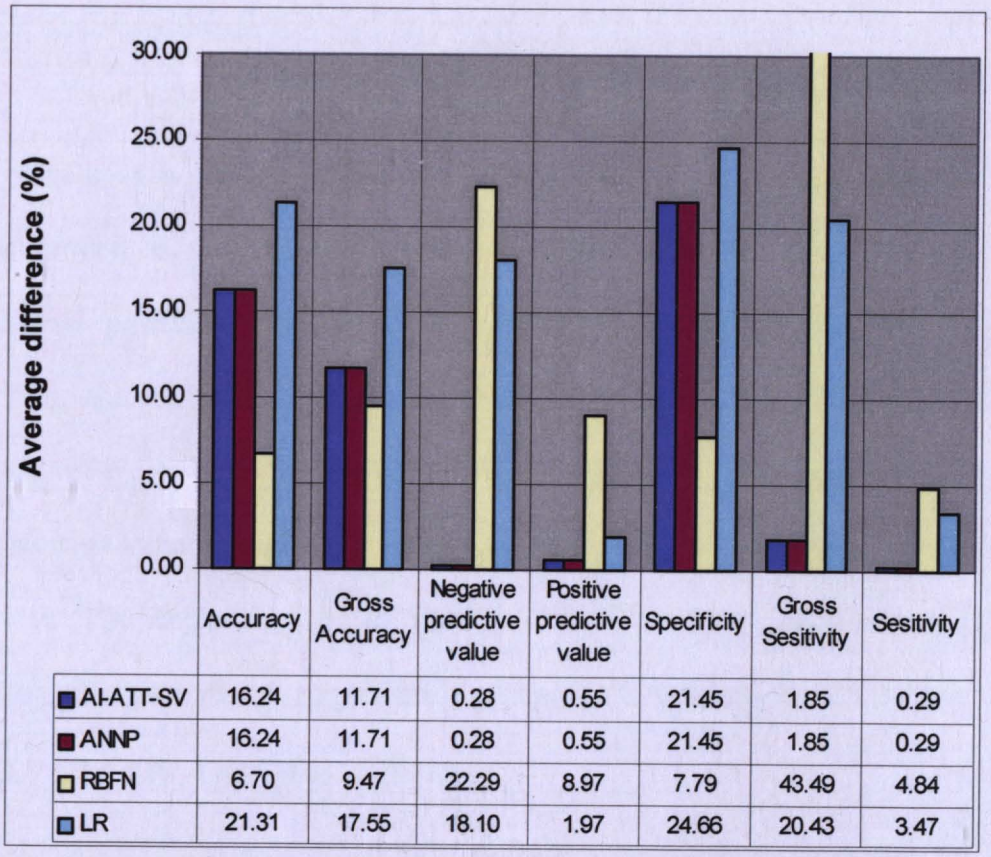


Fig.7.4 Methods comparison for screening variation prediction (including AI-ATT-SV). Balanced by SV classes.

The behaviour of the algorithm for the *Postal area data set* leads to similar results. Fig. 7.4 also shows the corresponding values.

If a close analysis is performed comparing the results obtained by the AI-ATT-SV with the best performing method for each option (with all models trained with the balanced by screening variation classes data, and assuming that the *attendance* prediction has been obtained with the AI-ATT algorithm), the following results are obtained (Fig. 7.5).



	Accuracy	Gross Accuracy	Negative predictive value	Positive predictive value	Specificity	Gross Sesitivity	Sesitivity
AI-ATT-SV	16.24	11.71	0.28	0.55	21.45	1.85	0.29
ANNP	16.24	11.71	0.28	0.55	21.45	1.85	0.29
RBFN	6.70	9.47	22.29	8.97	7.79	43.49	4.84
LR	21.31	17.55	18.10	1.97	24.66	20.43	3.47

Fig. 7.5 Comparison of the average differences per screening variation method with respect to the best performing model.

No single method can be selected as best performing with respect to all parameters. While for accuracy, gross accuracy and specificity, the RBFN is the better performing method, for the other parameters the ANNP (and consequently the AI-ATT-SV) are the best performing methods. As the final aim of the exercise is to detect those women likely to undergo screening variation, the ANNP and AI-ATT-SV are chosen as the preferred methods, since they not only achieve the best gross sensitivity and sensitivity in general (respective maxima of 92.20% and 15.84%, both achieved for the fourth episode *Townsend data set*), but even more, their predicted values (positive and negative) achieve the highest accuracy (27.79% and 89.02%, respectively). Furthermore, their gross accuracy deviates, on average, by only 2.24% from the RBFN's average results.

One more point needs to be taken into account when deciding which method is to be recommended to the Breast Screening Programme. As previously discussed for *attendance* prediction, the classification of a woman in relation to her current screening episode is also necessary before any of the discussed methods can be applied to the prediction of *screening variation*. The LR, ANNP and RBFN need this classification to be determined beforehand. The AI-ATT-SV performs it intrinsically. Therefore, the implementation of the AI-ATT-SV is recommended to the Breast Screening Programme for the prediction of *attendance* and *screening variation* of any given woman with a screening invitation.

Chapter 8

Blind Study of Attendance to the Breast Screening Programme

8.1 Experiment design

When introducing new methodologies into fields where conventional methods have traditionally been well established, it is a difficult task to instil conviction and belief into those methodologies. Disbelief and subjective thinking play an important role in the success or failure of a new technique.

In order to affirm the applicability of the method proposed for the prediction of attendance to a wide community, a blind study has been designed.

The Breast Screening Unit was requested to facilitate the study by supplying the invitation information for a batch of women invited after the last update of this study, i.e., data previously unseen. The unit would keep the attendance and screening results confidential. After the information has been processed and analysed by the predictive system, the results would then be sent back to the screening unit and, in return, they would release the actual attendance information of the batch for comparison with that predicted.

The batch sent by the screening unit consisted of 2169 women aged between 52 and 77 years at invitation and invited to be screened during September 2001. Of those, 226 were invited by the unit for the first time, while the rest (1943) had received at least one previous invitation to screening.

The final batch, after processing, contained 1814 women at different points in their screening history (from one to four screening episodes). Those women in the original batch who were in their fifth screening episode (355) were not included, since the system is not designed to predict attendance to the fifth episode.

8.2 Results

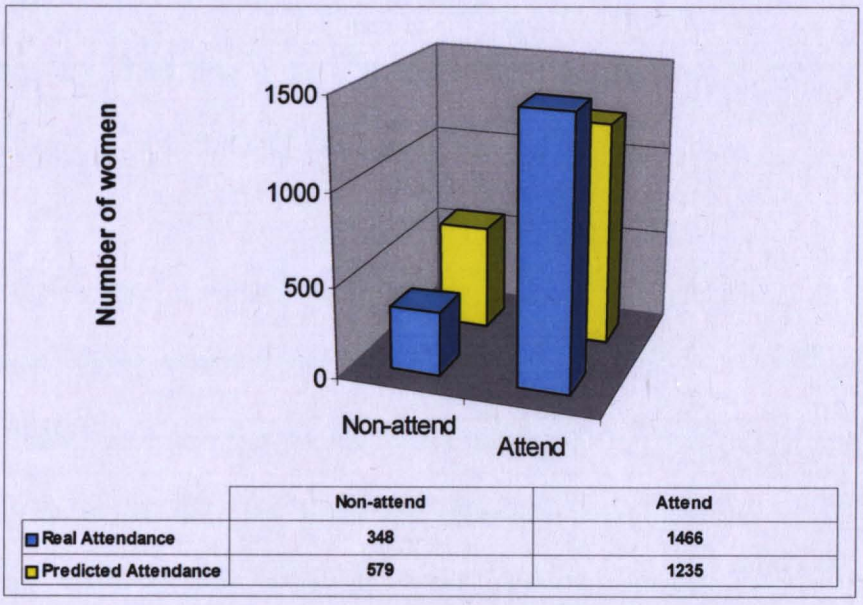
Table F.1 gives the detailed results from the final batch after processing.

Five hundred and seventy nine women out of 1814 were predicted not to attend the invitation. In reality, 348 women did not attend (Fig. 8.1). Moreover, from these 348 real non-attenders, the system correctly predicted 246 (70.69 %) of them.

Fig 8.2 shows the classification by prediction.

Both figures demonstrate that the predictive algorithm tends to predict more non-attendance than is actually the case. Even though this result may be construed as mathematically poor, clinically it can be useful. Since the ultimate aim of the screening programme is to identify those women likely not to attend in order to take appropriate

measures leading to their attendance, the identification of more women likely not to attend is more beneficial than the reverse.



	Non-attend	Attend
Real Attendance	348	1466
Predicted Attendance	579	1235

Fig. 8.1 Attendance prediction (blind study)

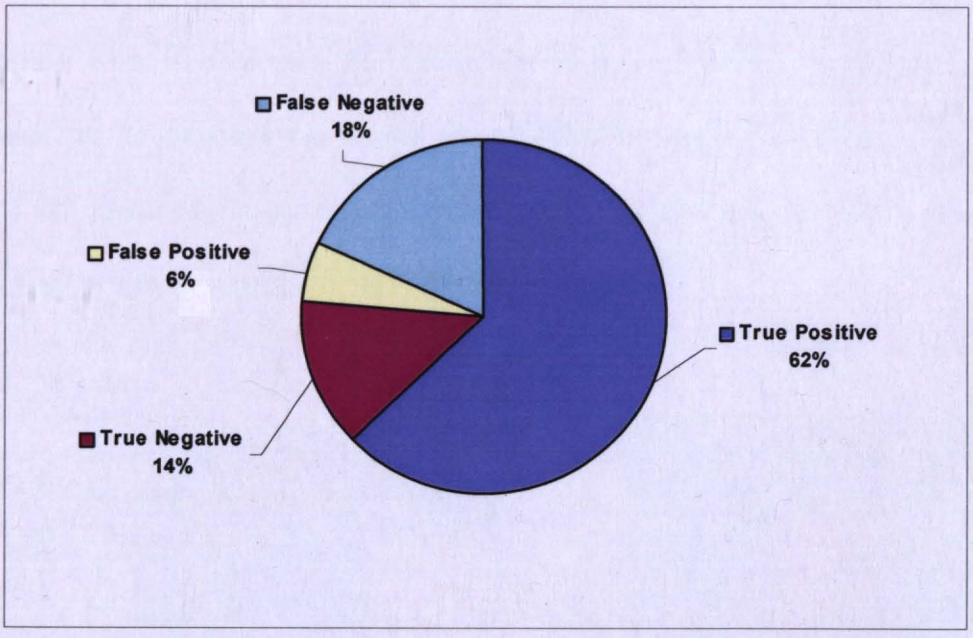


Fig. 8.2 Classification of prediction of attendance

The specificity of 70.69% obtained through this analysis implies that an equivalent percentage of women who are likely not to attend will be detected by the predictive system. The error in predicting the attendance of women who will not actually attend is reasonably low (6%), i.e., if the system predicts that the women will attend, there is a probability of 91.74% that they will do so.

Fig. 8.3 shows that, in terms of total accuracy, the system achieved approximately 76% (McNemar's Test p=0.0637) for the blind data study. These results are encouraging and give an initial tool for the prediction of not only the attendance of women to the screening invitation, but also, allow the identification of women who are likely not to attend. Once these women are identified, measures can be taken in order to increase their likelihood of attendance. Amongst such measures are the reissuing of letters of invitation and reminders, invitations through telephone calls, GP's invitations (instead of, or together with, the Breast Screening Programme invitation). Where a particular community or geographical area is identified as having a high percentage of women likely not to attend, measures can be taken, such as awareness campaigns and relocation of mobile screening units, amongst others.

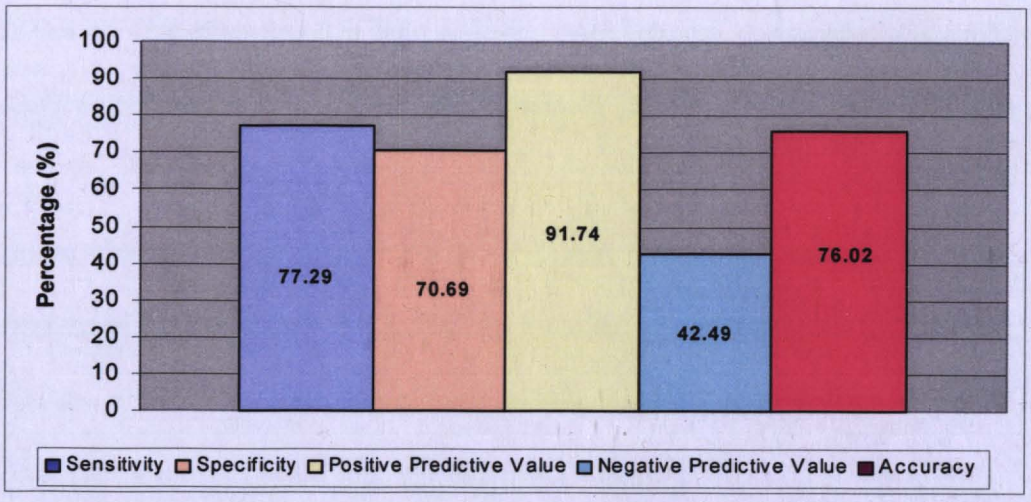


Fig. 8.3 Assessment of the attendance prediction for the blind study

Chapter 9

Conclusions and Recommendations

In this work, for the first time as far as can be ascertained, the entire database of screening history of the Solihull, Coventry and Warwickshire Breast Screening Unit has been submitted for analysis and modelling. Data on 137,051 non-symptomatic women invited for a preventive programme of breast screening has been analysed. Invitations covered the period from first offered appointments in 1989 until the closing year of the fourth screening round on 31st March 2000. Women were invited to take part in up to four different episodes, depending on their age when entering the programme.

A new analytical approach relating each screening episode to an individual woman as opposed to the three-yearly cycle imposed by the screening programme has been introduced. This approach has been followed with the aim of facilitating the use of the woman's previous history in future predictions of patient level features.

An extensive statistical analysis of the variables affecting *uptake* and *coverage* has been carried out. Section 4.4 has highlighted results pointing to the performance and achievements of the Unit. Two main points of weakness were detected, i) the attendance of the women at the screening invitation (81%); ii) the absence of a new

factor influencing the screening programme, named *screening variation. Screening variation* is defined as the time in days between the *date of first offered appointment* and actual *date of screening*. It can play an important role in the results of *coverage* as it has been observed that approximately 21% of screened women do not attend their screening on the date of initial offered appointment (although in the first 40 days after the date of invitation 98% of attending women had been screened).

In the light of these results three main recommendations have been suggested:

- an improved formulation of the current formula for *coverage* calculation, including the impact *screening variation* may have on such a parameter (defined in section 4.3.1);
- a reduction in the operational invitation period taking into account the effect of *screening variation* on the invitation scheduling (proposed in section 4.3.2);
- the introduction of a new performance measure defined as *Invitation rate,* designed to assess the capability of the system in reaching women (formulated in section 4.3.3).

An extensive analysis of the significance of those findings has been carried out in section 4.5. Following these results, and having highlighted the need for a predictive tool in order to identify those women who are likely not to attend or to change their screening invitations, efforts were focused on developing such predictive tools. Two main targets were set; first, the accurate prediction of *attendance* of any particular woman to the screening invitation; and second, predicting changes in dates of

appointments of any particular woman knowing that she will attend, i.e., predicting *screening variation* given the positive prediction of *attendance*.

An analysis of the predictors of *attendance* and *screening variation* was performed in section 5.1.3, and which represented the foundation for the features selection of inputs to the predictive methods.

Focusing on predictive techniques for *attendance* and *screening variation*, due to the structure of the invitation process (section 3.1.1), each invitation batch was noted to include women who were at different levels of their screening history (i.e. from those in their first invitation up to those in their fourth one). Another characteristic of the application was the existence of expected false "missing values", since even when a woman is invited, she does not always attend the invitation, and/or, even when having attended the initial invitation, she does not complete her present screening episode. This results in "missing" values for some variables, a problem likely to make some predictive algorithms fail or yield poor performance.

The performance of the *attendance* prediction models were assessed using the recognised concepts of sensitivity, specificity, positive and negative predictive values, and accuracy (introduced in section 5.2.3).

The AI models (i.e. ANNP and RBFN) achieved a better performance than the statistical model (LR) for *attendance*. This was an expected result given the typically poor manageability of missing values by this method and the particular data structure of the breast screening invitation process. Nevertheless, significant differences were

not detected between the performance of ANNP and RBFN which could point to the preference of one of them. This resulted in the development of the voting classifier used in Step 6 of the proposed AI-ATT algorithm.

The most desirable outcome for the screening units, in practical terms, is the detection of those rules that inform which women are likely not to attend, so that appropriate measures could be taken to improve their attendance. Applying rule induction techniques to the models generated by the ANN for each episode, and taking into account if the *Townsend deprivation score* information was available, decision trees for *attendance* and, in particular, for non-attendance have been identified in section 6.4. These results suggest that each episode and data set has different decision trees and rules for non-attendance based on the predictive method used. Although the screening unit could elect to apply those rules independently in order to target women likely not to attend the screening invitation, an ad-hoc computerised system could be used more sensibly to perform this task. Furthermore, the variety of methods and derived rules are not the only difficulties that have to be confronted when attempting to predict *attendance*. Almost all the generated models analysed incurred avoidable prediction errors, the most common being i) to predict positive *attendance* in cases when there was no invitation or ii) to make a general prediction when there was no *Townsend* information available. These errors need to be targeted and solved before any serious attempt is made to predict the *attendance* of women to the screening invitation.

The results obtained for the prediction of *attendance* using different techniques (ANN Pruning, RBF and Logistic Regression) were extensively analysed and an algorithm

for the prediction of *attendance* to the screening invitation (the AI-ATT algorithm) was proposed in section 6.5.1, and implemented under a Clementine environment.

The AI-ATT was concluded to yield a higher predictive performance in general than any of the other methods when all parameters were taken into account. Particularly, it achieved a better negative predictive value (non-attendance prediction) for all the episodes in both data sets (*Townsend* and *Postal Area*); the best (90.23%) was obtained in the second episode for the *Postal Area data set*. The AI-ATT algorithm may therefore be concluded to be a more robust alternative, as it achieves the most accurate results overall for all episodes, regardless of the differences in individual data sets for each episode. The LR method had the poorest performance for all episodes (only in the first one, when there is no previous screening history of the women, is it comparable with the other methods analysed).

The AI-ATT algorithm tended to increase its ability to identify non-attenders as the number of the episode increased, i.e., this showed the high potential of the algorithm to improve as more information of the screening history of the women became available.

The impact of those results on the Breast Screening Programme highlights that the proposed method lends the programme with a predictive tool able to reasonably predict women who are likely not to attend. Therefore, it gives the programme the opportunity of targeting such women efficiently in order to increase *attendance* in the first instance and decrease mortality due to breast cancer in the long term. Drawing up an exhaustive list of possible interventions with respect to non-attenders that could point to an improvement in *attendance* falls outside the scope of this study. However, the literature

points to the use of second appointment letters, personal phone calls, invitations through the women's GPs, localisation of screening vans nearer their homes and awareness, amongst other measures which have been proven to make an impact on women's attendance.

For those women attending the screening invitation, a desirable feature was to know the approximate date when they would attend the programme. In other words, it would be of benefit to predict if those women will undergo *screening variation* or not, and, if so, by what degree. Consequently, the development of a predictive algorithm for *screening variation* has been carried out and presented in chapter 7.

In order to assess the performance of the *screening variation* predictive models, new extensions of the concepts introduced in section 5.2.3 were defined in section 5.2.4, and two new concepts were introduced, *Gross sensitivity* and *Gross accuracy*.

The assessment of the models generated by the ANNP, RBFN and LR for screening variation prediction, together with the analysis of statistical predictors discussed in section 5.1, give results leading to the conclusion that, although the models developed fail to predict correctly the non-occurrence of *screening variation* or its accurate length, they were still able to predict occurrence of *screening variation* in general terms. At present, the use of such models is proposed as a tool for identifying women likely to incur *screening variation* so that further steps may be taken to improve their *attendance* on time, or as an indication of possible empty slots when planning screening sessions.

The AI-ATT-SV algorithm has been proposed in section 7.3.1 as an extension of the AI-ATT method for the prediction of *attendance* and *screening variation*. As a first approach to the prediction of *screening variation*, the model generated by the ANN Pruning was selected for use in this algorithm, given the stability of that model with respect to performance for all the episodes and data sets. However, this should not be taken as a final decision as previously discussed.

When more input information is available and a model is developed that is able to predict *screening variation* successfully, an analysis of the performance of the methods should be carried out and perhaps a voting method involving different models should be implemented in the corresponding step of the algorithm. Such work was not implemented in this project due to the unavailability of appropriate data to carry out an accurate prediction of *screening variation*. The Breast Screening Programme is strongly recommended to start collecting data relating to socio-economic factors of the women. Based on worldwide studies, particularly in the USA, data relating to educational background, income, ethnicity and religion, amongst other factors, have been shown to be closely correlated to attendance of women at screening and to the late presentation of symptomatic women. Therefore, such factors are quite likely to be strongly related to the occurrence of screening variation. Additionally, based on studies carried out in the UK, information should also be collected on factors such as distance of a woman's home to the screening facility (fixed or mobile unit). Other parameters that could be of interest to take into account are, i) information related to car ownership, and whether the woman drives to the appointment or uses public transport; ii) if the woman has grandchildren and is involved in looking after them while the parents work, or if she is still subject to working commitments. The setting up of a field study is suggested in the first instance,

in order to evaluate whether these parameters help in the prediction of screening variation in the Breast Screening Programme in the UK, and to estimate the extra cost that collecting such parameters would involve.

When introducing new methodologies into fields where conventional methods have been well established, convincing those who should use them is known to be intricate. Disbelief and subjective thinking play an important role in the success or failure of a new technique. Therefore, to demonstrate to the world community the validity of the proposed method for the prediction of *attendance*, a blind study was designed involving a batch of 2169 women screened during September 2001. The results discussed in Chapter 8 show that 70.7% of women who are likely not to attend were successfully identified by the predictive system, and that the error in predicting *attendance* in women who will not attend was only 6%. Also, if the system predicts that a woman will attend, there is a probability of 91.7% that the woman will do so. This encouraging result leads to propose to the Breast Screening Unit the use of the AI-ATT algorithm as a predictive tool for, not only the attendance of the women to the screening invitation, but also as a tool of identifying women who are likely not to attend. When those women are singled out, measures can be taken to increase their likelihood of attendance.

Results obtained in this project are based on data from only one screening unit. In order to make these methods available to other screening units, and to the whole screening programme, models with more geographically complete data need to be developed using the proposed algorithms.

This project takes a step towards increasing the attendance of women at screening based on individual prediction given previous history, and directs actions towards future participation. Although our approach has been validated, much more work is needed in order to make a real impact on the ultimate aim of the programme, that of reducing mortality due to breast cancer.

Finally, it is noteworthy to mention that most of the results obtained from this study have been published. Seven international conference papers have been presented: two in 2000 [54, 55], two in 2001 [56, 57], two in 2002 [58, 59], and one in 2003 [60]. In addition, two papers have been published in internal proceedings in Coventry University in 2001 and 2002 respectively [61, 62]. A full report [63] has been submitted and approved by the sponsoring collaborating body, the NHS Breast Screening Programme.

9.1 Contributions to the field

The work presented in this thesis makes a number of contributions to the field of Biomedical Computing.

Complete analysis of the screening unit's database

For the first time, since the foundation of the Solihull, Coventry and Warwickshire Breast Screening Unit, the entire screening database has been submitted to external scrutiny and analysed with the objective of modelling a women's behaviour towards screening. This is a major step forward in improving breast cancer screening outcomes in the region under analysis.

New analytical approach to understanding women's behaviour towards screening

In the present work, a new analytical approach to studying a women's behaviour towards screening is presented. Previously in the UK, any such analysis focused on the screening programme's structure rather than the individual woman. As a consequence, previous attempts to predict behaviour were only carried out at an aggregate level, making it impossible to detect individual women exhibiting behaviour likely to detract from the efficacy of the screening programme.

In our work, we approach the analysis from the point of view of the individual woman's history, facilitating the identification of those factors affecting most closely their behaviour towards screening, and prediction of future behaviour.

This is not only a step forward in the analysis of participation in this particular screening programme (i.e. the Breast Screening Programme), but also opens up possibilities for further studies of participation in other health screening programmes.

Improvement of the coverage formula

In this project, an evaluation of the actual *coverage* formula used by the Breast Cancer Screening Programme has been carried out and a new formula, including factors not previously taken into account, has been proposed.

Due to possible operational difficulties in applying the new *coverage* formula at national level within a short time, a reduction in the operational invitation period has been suggested, as an alternative.

Also, as part of the implementation of the above proposals, the use of the newly formulated *Invitation rate* is proposed as an auxiliary measurement in assessing the performance of the breast screening units and their effectiveness in reaching women (in the first place). This challenges existing practice, that currently focuses on the assessment of coverage by number of women screened, only.

Implementation of the proposed changes should, it is suggested, improve our ability to assess the effectiveness of the programme.

Identification of predictors of *attendance* and *screening variation* of the women invited for screening

A further contribution of this work is the identification of predictors of *attendance* and *screening variation.*

The main predictors of *attendance* identified for an individual woman to screening invitations (in order of importance) are:

- The postal area where the woman lives
- The Townsend deprivation score value for the woman's address
- The screening variation for the previous episode (i.e. if the woman changed the date of the invitation)
- The number of tests to which the woman was submitted in the previous screening episode

- The results obtained in previous screening episodes (including detection or not of cancer and occurrence or not of false positive results)
- The woman's age band at invitation
- The woman's attitude towards previous invitation (i.e. if she attended or not)

With respect to *screening variation* (change of date of screening appointments), the main predictors identified (in order of importance) are:

- The postal area where the woman lives
- The Townsend deprivation score value for the woman's address
- The results obtained in previous screening episodes
- The screening variation for the previous episode (i.e. if the woman changed the date of the invitation)
- The woman's age band at invitation

It should be noted that the predictors identified for *attendance*, agree with those reported in the literature. The identification of predictors for *screening variation* is reported for the first time in this study, although similar predictors are reported in the literature for related factors such as interval cancers and delayed presentation in symptomatic women.

Development of predictive algorithms for *attendance* and *screening variation*

Much of the novelty of this work lies in the development of algorithms that predict *attendance* (***AI-ATT***) and *screening variation* (***AI-ATT-SV***) for individual women. These algorithms are implemented using a novel approach, based on artificial intelligence methodology. Although this methodology has been successfully applied in other areas of cancer research, its use in the prediction of *attendance* to medical amenities, (in this case, the Breast Cancer Screening Programme) is new.

These algorithms constitute a valuable tool for the Programme, offering the potential to identify particular women likely not to attend or to change appointments such that appropriate measures may be taken to target the women or plan for changes in an optimal way.

The proposed algorithms have been validated not only in a laboratory situation, but also in real life screening practice in the unit under study.

9.2 Proposed future work

This project opens up a number of possibilities that could lead to improvement of the Breast Cancer Screening Programme in the first instance, and towards reducing mortality due to breast cancer in the longer term.

Areas recommended for future study include:

Improving the models achieved through identification of further predictors

As mentioned before, the analysis of socio-economic factors carried out in this project was subject to restrictions of data availability. Nevertheless, the literature reports other specific socio-demographic and economical factors that influence participation and retention in relation to medical amenities. Although most of these studies have been carried out elsewhere in the world, there is no proof that such factors will influence *attendance* or *screening variation* for a given screening unit, in this case. A field study is therefore required to verify the influence of such factors as a precursor to including them in future models.

Although the prediction of *attendance* also could benefit from such a study, it is likely to be of most benefit in improving the prediction for *screening* variation for which results were less accurate.

Improvements in the predictive algorithms

The accuracy of the AI-ATT predictive algorithm for the first episode was not as high as that achieved for subsequently episodes. There is a possibility that including a Logistic Regression model in the voting classifier for the first episode might improve results. This possibility should be investigated.

Furthermore, the existing algorithm would allow the inclusion of other Artificial Intelligence based models (e.g. Genetic Algorithms, Kohonen Networks, etc) within the voting classifier. It is possible that performance may be improved by extending the algorithm in this way.

Obtaining predictive models for further episodes

Due to the fact that women were only in their fourth screening episode when the data was made available to us, the algorithm lacks the capacity to predict the fifth and further episodes. It would be desirable to generate predictive models for such episodes. These could be accommodated within the AI-ATT and AI-ATT-SV algorithms respectively with little difficulty.

Generalisation to other units within the Breast Screening Programme

The methodology proposed in this project, although tested only for a single unit, could be easily extended for use in other units. Although, an initial study would be necessary to confirm that the predictors identified for this unit also apply to other units, the algorithm could be used with new generated predictive models trained with data specific to the domain.

A similar approach, using a well-distributed sample from all the screening units, could be applied to the whole country.

Generalisation to other screening programmes

Although designed for the Breast Cancer Screening Programme, the methodology proposed could also be extended to other screening programmes. Similar extensions to those recommended for its use in other units should be applied.

References

1. Quinn M., Babb P., Brock A., Kirby L., Jones J.: Cancer Trends in England and Wales 1950-1999. *National Statistics Studies on Medical and Population Subjects,* The Stationery Office. **66** (5) Breast, 40-45, 2001.
2. Mettlin C.: Global Breast Cancer Mortality Statistics. *CA Cancer Journal for Clinicians*. **49** (3), 138-144, 1999.
3. WHO Databank. WHO Mortality Database, United Kingdom: Breast (1950-1997). http://www-dep.iarc.fr, 2000.
4. Forrest, P.: Breast Cancer Screening, *Report to the Health Ministers of England, Wales, Scotland & Northern Ireland.* Her Majesty's Stationery Office, Department of Health and Social Security, 1986.
5. Blanks R.G., Moss S.M., McGahan C.E., Quinn M.J., Babb P.J.: Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *British Medical Journal.* **321** (7262), 665-669, 2000.
6. Stead M.J., Wallis M.G., Wheaton M.E.: Improving uptake in non-attenders of breast screening: selective use of second appointment. *Journal of Medical Screening.* **5** (2), 69-72, 1998.
7. Faux A., Lawrence G., Wheaton M., Wallis M., Jeffery C., Griffiths R.: Slippage in the NHS breast screening programme: an assessment of whether a three year screening round is being achieved. *Journal of Medical Screening.* **5** (2), 88-91, 1998.

8. Shapiro S., Coleman E.A., Broeders M., Codd M., De Koning H., Fracheboud J., Moss S., Paci E., Stachenko S., Ballard-Barbash R.: Review article. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. *International Journal of Epidemiology.* **27** (5), 735-742, 1988.

9. Colditz G.A., Hoaglin D.C., Berkey C.S.: Cancer incidence and Mortality: The priority of Screening Frequency and Population Coverage. *The Milbank Quarterly.* **75** (2), 147-173, 1997.

10. Ballard-Barbash R, Clabunde C., Paci E., Broeders M., Coleman E.A., Fracheboud J., Bouchard F., Rennert G. and Shapiro S.: Breast cancer screening in 21 countries: delivery of services, notification of results and outcomes ascertainment. *European Journal of Cancer Prevention.* **8** (5), 417-426, 1999.

11. Alexander F.E., Anderson T.J., Brown H.K., Forrest A.P., Hepburn W., Kirkpatrick A.E., McDonald C., Muir B.B., Prescott R.J., Shepherd S.M.: The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. *British Journal of Cancer.* **70**, 542-548, 1994.

12. Alexander F.E., Anderson T.J., Brown H.K., Forrest A.P.M., Hepburn W., Kirkpatrick A.E., Muir B.B., Prescott R.J., Smith A.: 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *The Lancet.* **353** (9168), 1999.

13. NHS Breast Screening Programme Review 1993.

14. NHS Breast Screening Programme Review 1994.

15. NHS Breast Screening Programme Review 1995.

16. NHS Breast Screening Programme Review 1997.

17. NHS Breast Screening Programme Review 1998.

18. Kerlikowske K., Grady D., Rubin S.M., Sandrock C., Ernster V.L.: Efficacy of screening mammography. A meta-analysis. *JAMA: The Journal of the American Medical Association.* 273, 149-154, 1995.

19. Mittra I.: Early detection of breast cancer in industrially developing countries. *Gan To Kagaku Ryoho: Japanese Journal of Cancer and Chemotherapy.* **22** (3), 230-235, 1995.

20. Watmough D.J., Bhargava S., Memon A., Syed F., Roy S., Sharma P.: For Debate Does breast cancer screening depend on a wobbly hypothesis? *Journal of Public Health Medicine.* **19** (4), 375-379, 1997.

21. Gotzsche P.C., Olsen O.: Is screening for breast cancer with mammography justifiable? *The Lancet.* **355** (9198), 2000.

22. Solomon L.J., Mickey R.M., Rairikar M.A., Worden J.K., Flynn B.S.: Three-Year Prospective Adherence to Three Breast Cancer Screening Modalities. *Preventive Medicine.* **27**, 781-786, 1998.

23. Dickersin K.: Breast screening in women aged 40-49 years: what next? *The Lancet.* **353** (9168), 1999.

24. Faux A.M., Richardson D.C., Lawrence G.M., Wheaton M.E., Wallis M.G.: Interval breast cancers in the NHS Breast Screening Programme: does the current definition exclude too many? *Journal of Medical Screening.* **4** (3), 169-173, 1997.

25. Lechner L., De Vries H., Offermans N.: Participation in a Breast Cancer Screening Program: Influence of Past Behaviour and Determinants on Future Screening Participation. *Preventive Medicine.* **26**, 473-482, 1997.

26. Blanks R.G., Moss S.M., Wallis M.G.: Use of two view mammography compared with one view in the detection of small invasive cancers: further results from the

National Health Service breast screening programme. *Journal of Medical Screening.* **4** (2), 98-101, 1997.

27. Blanks R.G., Wallis M.G., Moss S.M.: A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service breast screening programme. *Journal of Medical Screening.* **5** (4), 195-201, 1998.
28. Armitage P.: *Statistical Methods in Medical Research.* Blackwell Scientific Publications, 1971.
29. Puri B.K.: *Statistics for the health sciences: using SPSS.* Saunders, 1996.
30. Altman D.G.: *Practical Statistics for Medical Research.* Chapman & Hall, 1992.
31. Bland M.: *An introduction to Medical Statistics.* Oxford Medical Publications, 1987.
32. Patterson D.W.: *Artificial Neural Networks Theory and Applications.* Simon & Schuster (Asia) Pte Ltd, 1996.
33. Ripley B.D.: *Pattern Recognition and Neural Networks.* Cambridge University Press, 1996.
34. Berry M.J.A., Linoff G.: *Data Mining Techniques for Marketing, Sales, and Customer Support.* New York: John Wiley and Sons. 1997.
35. Bigus J.P.: *Data Mining with Neural Networks: Solving Business Problems—from Application Development to Decision Support.* New York: McGraw-Hill. 1996.
36. Bishop C.M.: *Neural Networks for Pattern Recognition.* Oxford: Oxford University Press. 1995.
37. Masters T.: *Practical Neural Network Recipes in C++.* Academic Press. 1993.
38. Masters T.: *Advanced Algorithms for Neural Networks: A C++ Sourcebook.* New York: John Wiley and Sons. 1995.

39. Goldberg D.E.: *Genetic Algorithms in Search, Optimization & Machine Learning.* Addison – Wesley Publishing Company, Inc., 1989.

40. Mitchell T.: *Machine Learning.* Boston: McGraw-Hill. 1997.

41. Quinlan R.: *C4.5: Programs for Machine Learning.* San Mateo: Morgan Kaufmann Publishers. 1993.

42. Collett D.: *Modelling binary data.* Chapman & Hall. 1991.

43. Agresti A.: *Categorical Data Analysis.* New York: John Wiley & Sons. 1990.

44. Agresti A.: *An Introduction to Categorical Data Analysis.* New York: John Wiley & Sons. 1996.

45. Cox D.R., Snell E.J.: *The Analysis of Binary Data.* 2nd ed. New York: John Wiley & Sons. 1989.

46. Hosmer D.W., Lemeshow S.: *Applied Logistic Regression.* New York: John Wiley & Sons. 1989.

47. McCullagh P., Nelder J.A.: *Generalized Linear Models.* 2nd ed. London: Chapman and Hall. 1989.

48. Seker H., Odetayo M.O., Petrovic D., Naguib R.N.G., Bartoli C., Alasio L., Lakshmi M.S., Sherbet G.V.: A fuzzy measurement-based assessment of breast cancer prognostic markers. *Proceedings of the IEEE-EMBS International Conference on Information Technology Applications in Biomedicine*, Washington, USA, 174-178, 2000.

49. Ji W., Naguib R.N.G., Ghoneim N.A.: Assessment of Bilharziasis History in Outcome Prediction of Bladder Cancer Using a Radial Basis Function Neural Network. *Proceedings of the IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, Washington DC, USA, 268-271, 2000.

50. Naguib R.N.G., Mat-Sakim H.A., Lakshmi M.S., Wadehra V., Lennard T.W.J., Bhatavdekar J., Sherbet G.V.: DNA Ploidy and Cell Cycle Distribution of Breast Cancer Aspirate Cells Measured by Image Cytometry and Analysed by Artificial Neural Networks for their Prognostic Significance. *IEEE Transactions on Information Technology in Biomedicine*. **3** (1), 61-69, 1999.

51. Seker H., Odetayo M., Petrovic D., Naguib R.N.G., Bartoli C., Alasio L., Lakshmi M.S., Sherbet G.V.: Assessment of Nodal Involvement and Survival Analysis in Breast Cancer Patients Using Image Cytometric Data: Statistical, Neural Network and Fuzzy Approaches. *Anticancer Research (International Journal of Cancer Research and Treatment)*. **22** (1), 433-438, 2002.

52. Seker H., Odetayo M., Petrovic D., Naguib R.N.G., Bartoli C., Alasio L., Lakshmi M.S., Sherbet G.V.: Prognostic Comparison of Statistical, Neural and Fuzzy Methods of Analysis of Breast Cancer Image Cytometric Data. *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Istanbul, Turkey, 2001.

53. Lancucki L., Babb P.J.: Monitoring and evaluating the performance of breast screening programmes: Use of the estimated cumulative attendance (ECA) as a measure of uptake. (*Personal Communication*).

54. Arochena H.E., Whitford A., Naguib R.N.G., Wheaton M.E., Wallis M.G.: Descriptive Analysis of Parameters Affecting a UK Breast Cancer Screening Unit. *Proceedings of the World Congress on Medical Physics and Biomedical Engineering*. Chicago, USA, 2000.

55. Arochena H.E., Whitford A., Naguib R.N.G., Mashhoudy H., Wheaton M.E., Wallis M.G.: Analysis of Attendance and Screening Variation in a UK Breast Cancer

Screening Unit. *Proceedings IEEE EMBS International Conference on Information Technology Applications in Biomedicine ITAB 2000*. Virginia, USA, 332-337, 2000.

56. Arochena H.E., Naguib R.N.G., Mashhoudy H., Wheaton M.E., Wallis M.G.: Screening Variation: Its Impact on Uptake and Coverage for the Breast Screening Programme. *Proceedings European Conference on Cancer Strategies & Outcomes. European Journal Of Cancer*. **37** (Supp. 2), S110, 2001.
57. Arochena H.E., Naguib R.N.G., Mashhoudy H., Wheaton M.E., Wallis M.G.: Factors Associated and Predictors of Attendance to the UK Breast Screening Programme. *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Istanbul, Turkey, 2001.
58. Arochena H.E., Todman A., Naguib R.N.G., Wheaton M.E., Wallis M.G.: Discussion of an Algorithm for the Prediction of Attendance at First Invitation from a UK NHS Breast Screening Unit. *Proceedings of the IEEE EMBS UK & RI Postgraduate Conference in Biomedical Engineering and Medical Physics*. IEEE EMB Student Society UK & RI. 17, 2002.
59. Arochena H.E., Naguib R.N.G., Todman A., Wheaton M.E., Wallis M.G.: An Algorithm for the Prediction of Attendance to a Breast Screening Unit in the UK. *Proceedings of the 2nd Joint EMBS-BMES Conference*. Houston, USA, 76-77, 2002.
60. Arochena H.E., Naguib R.N.G., Todman A., Wheaton M.E., Wallis M.G.: Blind Study for the Prediction of Attendance to a UK Breast Cancer Screening Unit. *Proceedings of the 4th Annual IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine ITAB 2003*. Birmingham, UK, 2003.
61. Arochena H.E., Naguib R.N.G., Mashhoudy H., Wheaton M.E., Wallis M.G.: Effect of Screening Variation on Uptake and Coverage Calculations for the Breast

Screening Programme in the UK. *Proceedings of the Symposium on Research in Computer Science.* Coventry University. Coventry, UK. 8-11, 2001.

62. Arochena H.E., Naguib R.N.G., Todman A., Wheaton M.E., Wallis M.G.: A Comparative Assessment of Methods Predicting Attendance to the Breast Screening Programme. *Proceedings of the 2nd Symposium on Research in Computer Science.* Coventry University. Coventry, UK. 21-26, 2002.
63. Arochena H.E., Naguib R.N.G., Mashhoudy H., Wheaton M.E., Wallis M.G.: Invitation and compliance of Breast Screening: A descriptive analysis of the first 10 years of data from the Warwickshire, Solihull and Coventry Breast Screening Unit. *Internal report to the Breast Screening Programme.* Coventry University, 2001.
64. UK Trial of Early Detection of Breast Cancer Group: 16-year mortality from breast cancer in the UK Trail of Early Detection of Breast Cancer. *The Lancet.* **353** (9168), 1909, 1999.
65. Olsen O., Gøtzsche P.C.: Cochrane review on screening for breast cancer with mammography. *The Lancet.* **358** (9290), 1340-1342, 2001.
66. Olsen O., Gøtzsche P.C.: Screening for breast cancer with mammography. *Cochrane Library*, **4**. Oxford:Update Software (in press).
67. Koning H.J.: Assessment of nationwide cancer-screening programmes. Commentary. *The Lancet*. **355** (9198), 2000.
68. National Institutes of Health (NIH) Consensus Development Panel: Breast cancer screening for women ages 40-49. *NIH Consensus Statement*. **15** (1), 1-35, 1997.
69. Jenkins C.N.H., McPhee S.J., Bird J.A., Pham G.Q., Nguyen B.H., Nguyen T., Lai K.Q., Wong C., Davis T.B.: Effect of a Media-Led Education Campaign on Breast and Cervical Cancer Screening among Vietnamese-American Women. *Preventive Medicine*. **28**, 395-406, 1999.

70. Paskett E.D., McMahon K., Tatum C., Velez R., Shelton B., Case L.D., Wofford J., Moran W., Wymer A.: Clinic-Based Interventions to Promote Breast and Cervical Cancer Screening. *Preventive Medicine*. **27**, 120-128, 1998.

71. Giorgi D., Giordano L., Senore C., Merlino G., Negri R., Cancian M., Lerda M., Segnan N., Del Turco M.R.: General practitioners and mammographic screening uptake: influence of different modalities of general practitioner participation. Working Group. *Tumori.* **86** (2), 124-129, 2000.

72. Lagerlund M., Sparén P., Thurfjell E., Ekbom A., Lambe M.: Predictors of non-attendance in a population-based mammography screening programme: socio-demographic factors and aspects of health behaviour. *European Journal of Cancer Prevention*. **9** (1), 25-33, 2000.

73. Martinez L.J.: *The role of physics testing in breast cancer screening*. Ph.D. Thesis. City University. 1995.

74. Chen H.H., Duffy S.W., Tabar L.: A Markov chain method to estimate the tumor progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *Statistician*. **45** (3), 307-317, 1996.

75. Chen H.H., Thurfjell E., Duffy S.W., Tabar L.: Evaluation by Markov chain models of a non-randomised breast cancer screening programme in women aged under 50 years in Sweden. *Journal of Epidemiology and Community Health*. **52** (5), 329-335, 1998.

76. Williams E.M.I.: *Breast cancer screening: studies in the evaluation of a district service*. M.D. Thesis. London King's College Hospital Medical School. 1989.

77. Guidelines on Quality Assurance Visit. NHS Breast Screening Programme. 1998.

78. Muir Gray J.A.: *A Draft Set of Criteria for Evaluation and Quality Assurance. Screening Publications*. Hall The Printer, Oxford. 1990.

79. KC63 Statistics Table Definitions. FHS Computer Agency, Version 3, 1994.

80. Breast Screening Programme, England:1999-2000. Bulletin 2001/10. March 2000. http://www.doh.gov.uk/public/sb0110.htm, Crown Copyright. 2001.

81. Richards M.A., Westcombe A.M., Love S.B., Littlejohns P., Ramirez A.J.: Influence of delay on survival in patients with breast cancer: a systematic review. *The Lancet.* **353** (9159), 1119, 1999.

82. Edwards N.I., Jones D.A.: Uptake of breast cancer screening in older women. *Age & Ageing.* **29** (2), 131-135, 2000.

83. Bell T.S., Branston L.K., Newcombe R.G., Barton, G.R.: Interventions to improve uptake of breast screening in inner city Cardiff general practices with ethnic minority lists. *Ethnicity & Health.* **4** (4), 277-284, 1999.

84. Gatrell A., Garnett S., Rigby J., Maddocks A., Kirwan M.: Uptake of screening for breast cancer in south Lancashire. *Public Health.* **112** (5), 297-301, 1998.

85. Thompson D.L.: *The National Health Service breast screening programme in Sheffield: service delivery and uptake*. Ph.D. Thesis. Sheffield University. 1998.

86. Bish A.M.: *Predicting and explaining uptake of cervical screening: the role of social cognitions*. Ph.D. Thesis City University. 1997.

87. Gatrell A., Garnett S., Rigby J., Maddocks A., Kirwan M.: Uptake of screening for breast cancer in South Lancashire. *Public Health.* **112**, 297-301, 1998.

88. Challier B., Meslans Y., Viel J.F.: Deprived areas and attendance to screening of cervix uteri cancer in a French region. *Cancer Causes Control.* **11** (2), 157-162, 2002.

89. Frazier E.L., Jiles R.B., Mayberry, R.: Use of Screening Mammography and Clinical Breast Examinations among Black, Hispanic, and White Women. *Preventive Medicine*. **25**, 118-125, 1996.

90. Ong G.J.: *The needs and experiences of women who are recalled for further investigation of breast screening*. D.Phil. Thesis. Oxford University. 1997.

91. Rajaram S.S., Rashidi A.: Minority Women and Breast Cancer Screening: The Role of Cultural Explanatory Models. *Preventive Medicine*. **27**, 757-764, 1998.

92. Maxwell A.J.: Relocation of a static breast screening unit: a study of factors affecting attendance. *Journal of Medical Screening*. **7**(2), 114-115, 2000.

93. Jullien J.A., Zakrzewska J.M., Downer M.C., Speight P.M.: Attendance and compliance at an oral cancer screening programme in a general medical practice. *European Journal Cancer. B. Oral Oncology*. **31B** (3). 202-206, 1995.

94. Townsend P., Phillimore P., Beattie A.: *Health and deprivation: inequality and the north.* London, Croom Helm. 1988.

95. Payne G., Payne J., Hyde M.: 'Refuse of All Classes'? Social Indicators and Social Deprivation. *Sociological Research Online*. **1** (1). http://www.socresonline.org.uk/1/1/3.html. 1996.

96. Collins S.E., Haining R.P., Bowns I.R., Crofts D.J., Williams T.S., Rigby A.S., Hall D.M.B.: Errors in postcode to enumeration district mapping and their effect on small area analyses of health data. *The Journal of Public Health Medicine*. **20** (3), 1998.

97. Ramirez A.J., Westcombe A.M., Burgess C.C., Sutton S., Littlejohns P., Richards M.A.: Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. *The Lancet*. **353** (9159), 1127, 1999.

98. Faux A.M.J.: *Breast cancer in the West Midlands: an evaluation of screening, treatment and survival*. Ph.D. Thesis. Birmingham University. 1997.

99. Threlfall A.G., Woodman C.B.J., Prior, P.: Breast screening programme: should the interval between test depend on age? *The Lancet*. **349** (9050), 472, 1997.

100. McDonald M.: *Breast Screening and the consequences of recall for further assessment.* Ph.D. Thesis. Huddersfield University. 1998.

101. Ajzen I.: The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes.* **50,** 179-211, 1991.

102. Ripley R.M.: *Neural networks for breast cancer prognosis*. D.Phil. Thesis. Oxford University. 1998.

103. Kappen H.J., Neijt J.P.: Advanced ovarian cancer. Neural network analysis to predict treatment outcome. *Annals of Oncology.* **4** (4), S31-34, 1993.

104. Ravdin P.M., Clark G.M.: A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*. **22** (3), 285-93, 1992.

105. Burke H.B, Rosen D.B., Goodman P.H.: Comparing the prediction accuracy of statistical models and artificial neural networks in breast cancer. *NIPS*94 Post-Conference Workshop: Neural Network Applications in Medicine*. Vail, CO, USA, 1994.

106. Edwards J.G., Faux S.P., Plummer S.M., Abrams K.R., Walker R.A., Waller D.A., O'Byrne K.J.: Cyclooxygenase-2 expression is a novel prognostic factor in malignant mesothelioma. *Clinical Cancer Research.* **8** (6), 1857-1862, 2002.

107. Theeuwen M., Kappen B., Neijt J.: Neural network analysis to predict treatment outcome in patients with ovarian cancer. http://servius.mbfys.kun.nl/mbfys/people/bert/. 1995.

108. Kappen H.J., Neijt J.P.: Neural network analysis to predict treatment outcome. *Annals of Oncology.* **4** (Suppl. 4), S31-S34, 1993.

109. Ponsan C.: *Computing with words for data mining*. Ph.D. Thesis. Bristol University. 2000.

110. Hammond P., Speight P.M.: Screening for Risk of Oral Cancer and Pre-cancer. http://www.aim.unipv.it/pub/papers/idamap98/hammond.ps.gz. 1998.

111. Kent S.: *Diagnosis of Oral Cancer using genetic Programming*. Technical Report CNES-96-04. Department of Computer Science and Information Systems. Centre for Neural and Evolutionary Systems (CNES). Brunel University. UK. 1996.

112. Burke H.B.: The importance of Artificial Neural Networks in Biomedicine. *Applications of Neural Networks in Environment, Energy, and Health*. World Scientific Publishing, Singapore, Chp. 17, 145-153, 1996.

113. Harbeck N., Kates R., Ulm K., Graeff H., Schmitt M.: Neural network analysis of follow-up data in primary breast cancer. *The International Journal of Biological Markers*. **15** (1), 116-122, 2000.

114. Cooper R.A., Weeks A.J.: *Data, Models and Statistical Analysis*. Philip Allan Publishers Limited, 1983.

115. Everitt B.: *The Analysis of Contingency Tables*. Chapman and Hall, 1977.

116. Schwarzer G., Vach W., Schumacher M.: On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Statistics in Medicine*. **19**, 541-561, 2000.

117. Wilson C.: *Coverage and Uptake in the NHS Breast Screening Program*. MSc Thesis, Coventry University, 1999.

118. Microsoft Excel 97 SR-1. On-line Help. 1997.

119. West Midlands Cancer Intelligence Unit Data. (Personal communication); 2000.

120. SPSS for Windows 9.0.0 : *On-line help*. SPSS Inc. 1989-1999.

121. Clementine Data Mining System: *User Guide*. Integral Solutions Limited. 1998.

122. Kuncheva L.I.: *Fuzzy Classifier Design*. Physica-Verlag Heidelberg, New York, 2000.

123. Jakubowski J., Kwiatos K., Chwaleba A., Osowski S.: Higher Order Statistics and Neural Network for Tremor Recognition. *IEEE Transactions on Biomedical Engineering*. **49** (2), 152-159, 2002.

124. Intrator O., Intrator N.: Interpreting Neural-Network Results: a Simulation Study. http://www.math.tau.ac.il/~nin/papers/interpcsda.pdf . To appear on: *Computational Statistics and Data Analysis*. 2001.

125. Abdul-Kareem S., Baba S., Zubairi Y.Z., Prasad U., Wahid M.I.A.: Back Propagation Neural Network For Medical Prognosis: A Comparison Of Different Training Algorithms. http://www.sat.ait.ac.th/ej-sat/articles/3.1/sarmeem.pdf . 2001.

126. Attikiouzel Y., DeSilva C.J.S.: Applications Of Neural Networks In Medicine. *Australasian Physical & Engineering Sciences in Medicine*. **18** (3), 158-164, 1995.

127. Rogers S.K., Ruck D.W., Kabrisky M.: Artificial neural networks for early detection and diagnosis of cancer. *Cancer Letters*. **77**, 79-83, 1994.

128. Luenberger D.G.: *Introduction to linear and non-linear programming*. Editorial Pueblo y Educación, 1984.

129. Thimm G., Fiesler E.: Neural Network Pruning and Pruning Parameters. http://www.idiap.ch/neural/network.html. 1996.

130. Orr M. J. L.: Introduction to Radial Basis Function Networks. http://www.anc.ed.ac.uk/~mjo/intro/intro.html. 1996.